The Dissertation Committee for Xiaojing Ma Certifies that this is the approved version of the following dissertation:

Studies on HIV-1 nucleocapsid chaperone role in protein/nucleic acid interactions by single molecule spectroscopy approaches

Committee:

Paul F. Barbara, Supervisor

Brian A. Korgel

David A. Vanden Bout

Graeme A. Henkelman

Rick Russell

Studies on HIV-1 nucleocapsid chaperone role in protein/nucleic acid interactions by single molecule spectroscopy approaches

by

Xiaojing Ma, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2009

Dedication

我将这本毕业论文献给我的父亲母亲

Acknowledgements

The author thanks everyone who has had some input into the completion of this dissertation. In particular, I acknowledge my PhD advisor, Professor Paul F. Barbara for his supervision and support through my graduate studies for more than four years. I couldn't ask for a better place than Dr. Barbara's group to pursue my PhD degree. I have learned a lot from him not only about the science, but more importantly, how to be a great scientist. His endless passion about science and courage of facing various challenges during the journey of pursing the truth has set a high standard for my future career.

I am also very grateful to my colleagues and some former group members who are involved in the HIV project, Dr. Hsiao-wei Liu, Dr. Yining Zeng, Dr. Hui Wang, Dr. Yongjin Zhu, Dr. Christy Landes and Yeon-Joo Kim, without whom the work would not be completed gracefully.

From the bottom of my heart, I thank all the people who helped me prepared for the next step in my career. Most importantly, I thank my families for their endless support and love.

Studies on HIV-1 nucleocapsid chaperone role in protein/nucleic acid interactions by single molecule spectroscopy approaches

Publication No._____

Xiaojing Ma, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Paul F. Barbara

HIV-NC is a multifunctional protein which plays an important role in almost every step of the retroviral life cycle. NC is essential in catalyzing stand transfers of HIV-1 reverse transcription, including the annealing of the transactivation response element (TAR) of the viral genome to the complementary TAR DNA in minus-strong-stop DNA. In this dissertation, the research starts with focus on elucidating the reaction mechanism of NC-facilitated TAR DNA/RNA annealing using single molecule spectroscopy (SMS) approaches. The results indicate that nucleation of TAR DNA/RNA annealing occurs in an encounter complex form in which one or two DNA/RNA strands in the partially open "Y" form associated with multiple NC molecules. This encounter complex leads to annealing through the 3'/5' termini, namely "zipper" pathway and the annealing through the hairpin loop region, namely "kissing" pathway. By employing target oligonucleotides for specific TAR regions, we directly probed kinetic reversibility and the chaperone role of NC. Concentration-dependence of NC chaperoned melting and annealing of TAR

hairpins was investigated and the results further support the proposed reaction mechanism. Additionally, we used a single-stranded DNA (ssDNA) as model to study ssDNA conformational change upon NC binding. Here we present observation of NC binding to $d(TG)_n$ and $d(T)_n$, including NC effect on flexibility and conformation of these oligonucleotides chains. Our results reveal that the rigidity of ssDNA chain is dramatically reduced through interaction with NC. Meanwhile the results of NC dissociation experiments indicate the interaction of NC/ssDNA is complex and heterogeneous. Finally, we used SMS *in vitro* to systematically compare and contrast the RNA/protein interactions for the zinc-finger-binding-motif protein (NC) and the arginine-rich-binding-motif (ARM) protein (Tat) encoded by HIV-1. Tat and NC use different RNA binding motifs to recognize and interact with RNA hairpin, giving rise to very different changes in the RNA secondary structure upon protein binding. Competition experiments show that the presence of Tat can effectively inhibit the NC binding-induced local melting of TAR RNA hairpins. These results indicate that Tat specifically binds and stabilizes the TAR RNA hairpin structure, which likely inhibits the local melting of the hairpin induced by NC.

Table of Contents

Chapter 1: Introduction and Dissertation Overview

Studies of viral protein at the molecular level have provided insights into the fundamental cellular mechanisms. Intensive studies on the human immunodeficiency virus type 1 (HIV-1) have provided many examples of this process with new information about viral replication, assembly and nucleic acid/protein interaction (1-3). Study of how the protein involved in the biomolecule reaction using novel single molecule spectroscopic approaches opens new opportunity to understand detailed reaction mechanism and kinetics (4-11). In this dissertation, one important HIV-1 viral proteins, nucleocapsid protein (NC) including its role in the virus replication cycle, in particular, NC chaperone function in nucleic acid structural rearrangement will be discussed.

STRUCTURE OF THE HIV-1 VIRION

The HIV-1 virus is a retrovirus consisting of two copies of single-stranded RNA as genome encapsulated by viral structural proteins (3). The virus is enveloped by a lipid bilayer, which is derived from the host cell membrane. Inside the envelope is a matrix shell containing about 2,000 copies of the matrix protein (MA). A cone-shaped capsid composing of about 2,000 copies of capsid protein (CA) is located in the center of the virus. The capsid particle contains two copies of the viral genome in the form of RNA which are stabilized as a ribonucleoprotein complex with ~2,000 copies of nucleocapsid protein (NC) (2, 3). Within the complex, there are three essential virally encoded enzymes: reverse transcriptase (RT) and integrase (IN) and protease (PR). Some accessory proteins, such as Vif, Vpr and Nef are also packaged within the virus particle, while Tat and Rev are not packed in the virion (3).

OVERVIEW OF HIV-1 VIRUS LIFE CYCLE

The replication process of HIV-1 can be divided into two phases (1, 3). The early phase begins with the recognition of the target cell followed by the viral material entering the cell. Reverse transcription occurs and the viral genome is reverse transcribed into DNA, which is later transported into the nucleus and integrated into a host cell genome. The late phase begins with regulated expression of the integrated proviral DNA, and involves all the genome expression up to the virus budding and maturation.

The HIV-1 virus binds specifically to cells bearing CD4, a protein that functions in immune recognition. Subsequent to membrane fusion, the virion is uncoated and all its genome and viral proteins are released into host cell cytoplasm. Reverse transcription is initiated by annealing cellular $tRNA^{Lys}$ to the primer binding site on the viral RNA genome. The single-stranded viral RNA is reverse transcribed into double-stranded DNA which is then transported into cellular nucleus and integrated into the host cell genome by IN. Spliced and unspliced mRNA are synthesized in the late phase of virus life cycle. The mRNA transcripts are transported out of the nucleus for translation. Initially, short spliced mRNA encoded for the regulatory proteins, such as Tat and Rev are synthesized. Unspliced and singly spliced viral mRNAs are expressed to produce Gag and Gag-Pol polyproteins as well as other accessory viral proteins. Immature viral particle encapsulated two copies of the unspliced viral RNA and Gag bud out from the cell membrane. Subsequent to budding, the polyproteins are cleaved by PR to produce other enzymes as well as structural proteins including NC (1, 3).

THE HIV-1 NUCLEOCAPSID PROTEIN (NC)

HIV-1 NC is 55 amino acids in length and has two highly conserved CCHC-type zinc fingers, each of which binds a zinc ion (2)(Fig. 1.1). NC is a multifunctional protein which plays a role in almost every step of the retroviral life cycle, from reverse transcription and DNA integration to packaging and assembly (2). While some NC functions, such as genomic RNA packaging, are believed to involve sequence-specific binding to nucleic acids, NC also displays more general, nonsequence-specific nucleic-acid binding properties (2). In addition to its role as a structural protein that stabilizes the virion, NC also serves as a nucleic acid chaperone that catalyzes the rearrangement of both DNA and RNA into thermodynamically more stable structures and promotes several strand annealing reactions during reverse transcription (12-21) (Fig. 1.2). Transactivation response element (TAR) contains the first 59 nucleotides of the R region within strong-stop DNA which folds into a stable stem-loop structure (2). NC is essential to facilitate the strand transfer step by rearranging the structure of both TAR RNA and TAR DNA and promoting the hybridization of TAR DNA with TAR RNA located in the 3' end of viral genome (Fig. 2, step 3) (22, 23).

The chaperone activity of NC is believed to arise from two main consequences of NC binding to nucleic acids. First, NC lowers the energy barrier for annealing by partially melting the Watson–Crick pairing of the duplex regions due to NC preference for binding to single-stranded bases of both RNA and DNA (22, 24, 25). Second, NC lowers the energy cost of bringing two complementary hairpins together to form encounter complexes by screening the negative charges of the hairpins (23, 26-29).

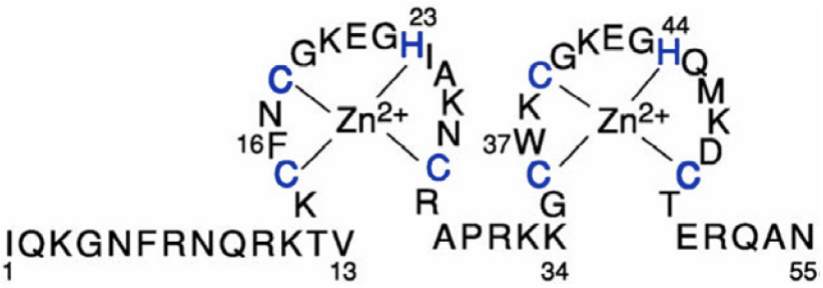


Fig. 1.1 Sequence of NC protein used in this study.

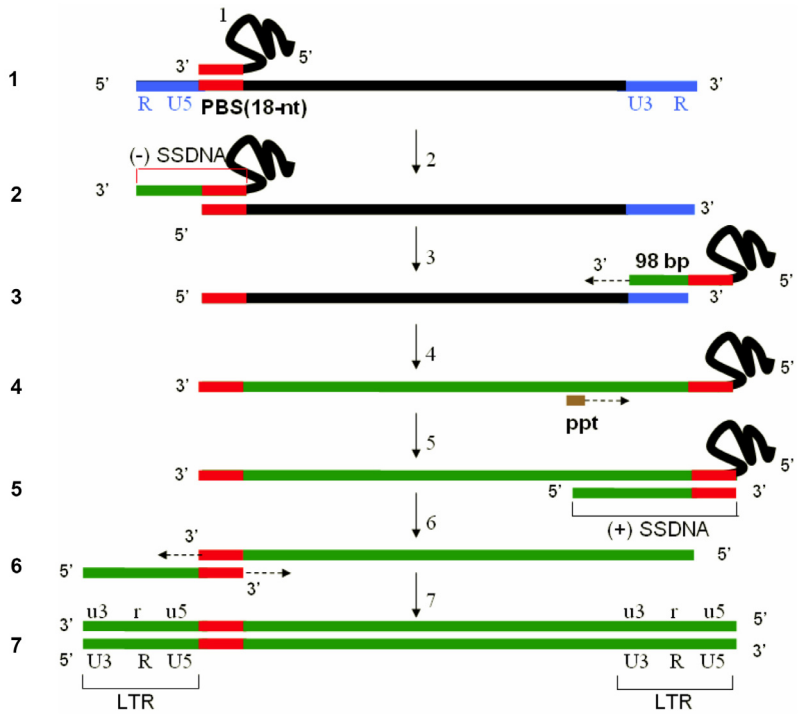


Fig. 1.2 Schematic illustration of HIV-1 reverse transcription.

Step 1: Annealing of $tRNA^{Lys,3}$ to the PBS. Step 2: Synthesis of (-) SSDNA and digest of RNA template by RT. Step 3: Minus-strand transfer. Step 4: Elongation of minus-strand DNA to the 3' end. Step 5: Synthesis of (+) SSDNA initiated from PPT as a primer. Step 6: Plus strand transfer. Step 7: Elongation of both DNA strands. Abbreviations: U5/3, 5'/3' untranslated region; PBS, primer binding site; SSDNA, strong stop DNA; PPT, polypurine tract.

SINGLE-MOLECULE SPECTROSCOPY AND ITS APPLICATION IN STUDIES OF NC

The greatest strength of single-molecule spectroscopy (SMS) is its ability to directly unravel molecule structures, dynamics and kinetics in highly heterogeneous and complex systems like biological complexes (7, 11, 30, 31). SMS has been successfully applied to investigate structural dynamics of nucleic acids, protein-protein interactions (32, 33), nucleic acids/protein interactions *in vitro* (12, 13, 34-36). New spectroscopic tools and sample preparation methods have been developed to tackle complicated biological questions during recent decades (37-39). The results have been led to direct investigations on some molecular properties that would be hidden within ensemble averaging process, such as previously unknown reaction pathways, intermediates and conformational dynamics.

Single-molecule Fluorescence Resonance Energy Transfer (SM-FRET) is one of the most popular techniques applied in biophysical studies (40). These single-molecule spectroscopic approaches allow one to observe one molecule at a time, and resolve subpopulations and heterogeneity in the biological systems. SMS approaches also reveal transient intermediates and temporal heterogeneous by recording individual transient in real time. These abilities to reveal intricate reaction mechanism makes SMS a unique and powerful tool used extensively by researchers.

The effect of NC binding on the conformations of TAR DNA haipins has been successfully investigated by single-molecule spectroscopic approach (12, 13). The data clearly demonstrates that NC shifts the equilibrium secondary structure of TAR DNA hairpins from a fully “closed” conformation to a “partially open” conformation. In order to explore the detailed reaction mechanism and kinetics of NC chaperoned nucleic acids structural rearrangement, we applied single-molecule fluorescence resonance energy transfer (discussed in the following section) as a main approach in my dissertation work.

SINGLE-MOLECULE FLUORESCENCE RESONANCE ENERGY TRANSFER SPECTROSCOPY (SM-FRET)

Fluorescence resonance energy transfer (FRET) can be used as a spectroscopic ruler because of its strong dependence on distance between two dye molecules (40). A small change in distance between two sites of molecules where two dyes are attached can be reflected in a change in FRET. Therefore, the FRET approach is attractive to be applied in investigation of biological molecules, their relative motions, dynamics and interaction between various molecules (41). FRET detected at the single-molecule level provides new opportunities to probe the structural changes and detailed kinetics of the complex biological molecules. Since the first demonstration of single-molecule FRET (42), many successful examples have provided new insights to the biological systems, meanwhile the methodology of SM-FRET have been developed and improved along the application of this technique.

The principle of FRET is the energy transfer in a non-radiative manner from a donor chromophore (dye) to an accepter chromophore in distance ranging from 10 to 75 Å (33, 43). FRET occurs through long range dipole-dipole interaction between a pair of chromophores. After absorption of a photon, the donor chromophore may relax its excess energy in a number of pathways, including fluorescence (k_F), quenching (k_Q), crossing to a triplet state (k_{ISC}), vibrational relaxation (k_{IC}), or non-radiative energy transfer. The fluorescence life time of donor dye depends on the sum of all the possible relaxation pathways.

$$\tau_D^{-1} = k_F + k_{ISC} + k_{IC} + k_Q \qquad (1)$$

If a suitable acceptor dye is present, the dipole-dipole interaction will results in an additional term k_{FRET}.

$$\tau_D^{-1} = k_F + k_{ISC} + k_{IC} + k_Q + k_{FRET} \qquad (2)$$

The energy transfer efficiency depends on the dye pair properties as well as the surrounding environment.

$$k_{FRET} = \frac{9000(\ln 10)\kappa^2 Q_D J}{128\pi^5 n^4 N_A \tau_D R^6} \qquad (3)$$

FRET efficiency is a function of refractive index of the medium between the donor/acceptor dyes *n*, the fluorescence life time τ_D and the quantum yield Q_n of donor dye in the absence of acceptor dyes, the normalized spectra overlap integral *J* [$M^{-1}cm^3$], the separation between two dye molecules *R*[cm], the Avogadro's number N_A and the orientation factor κ . It is often assumed that the dipole moments of two dyes molecules can freely rotate in all directions; in this case, κ^2 =2/3. The key idea for many FRET experiments is the relation between energy transfer efficiency and the distance between the donor and acceptor dye molecules. The equation can be simplified as the following

$$k_{FRET} = \frac{1}{\tau_D}\left(\frac{R_0}{R}\right)^6 \text{ , where } R_0 = \frac{9000(\ln 10)\kappa^2 Q_D J}{128\pi^5 n^4 N_A} \qquad (4)$$

R_0 is the distance at which 50% of the excitation energy is transferred to acceptor and is called Förster distance. The simple form is often used as R_0 effectively defines the relationship for a pair of donor/acceptor dyes. For common experiments, the parameters defining R_0 change little. However, control experiments and care must be taken before drawing the conclusion based only on observed FRET efficiency changes. In particular, quantum yield of the dyes can change with changing of solvent/ buffer solutions (43).

In many FRET experiments, E_{FRET} is the parameter need to be calculated. It is defined as the ratio of energy transfer rate to the sum of all the donor relaxation rates.

$$E_{FRET} = \frac{k_{FRET}}{k_F + k_{ISC} + k_{IC} + k_Q + k_{FRET}} \qquad (5)$$

Thus, we can calculate E_{FRET} from

$$E_{FRET} = \frac{1}{1+(\frac{R}{R_0})^6} \qquad (6)$$

The energy transfer efficiency changes with six power of the distance between two dye molecules therefore it can be used as a powerful molecular probe at the nanometer length scale. One can easily imagine that the structural change of biological molecules or interaction between two different molecules can be detected by changes in E_{FRET}.

Experimentally E_{FRET} can be calculated using the relative fluorescence intensities

$$E_{FRET} = \frac{I_A}{I_A + (\frac{\Phi_A \eta_A}{\Phi_D \eta_D}) I_D} \qquad (7)$$

Where, I_D and I_A are the fluorescence intensities of the donor and acceptor dyes respectively, Φ_D and Φ_A are the quantum yields of the donor and acceptor molecules, η_D and η_A are the detection efficiencies of the experiments for two dye molecules. Under our experimental condition, the ratio of $\Phi_A * \eta_A / \Phi_D * \eta_D$ set to be unity. Therefore, the apparent $E_A(t)$ given by the ratio of acceptor intensity to the sum of acceptor plus donor intensity is equal to $E_{FRET}(t)$ as shown in the following equation

$$E_A = \frac{I_A}{I_A + I_D} \qquad (8)$$

FRET values can be determined from each individual molecules in the way described above and their histogram can directly give information about FRET value distribution. In our experimental setup (scanning confocal microscopy), we use a combination of *image scanning mode* and *individual trajectory mode* to time-resolve the structural dynamics of the target biomolecules at multiple time scales. We use the *image scanning mode* to globally image the samples and dynamics on the time scale of minutes

or longer, and use the *individual trajectory mode* to focus on one particular molecule at a time in order to study dynamics on the time scales from milliseconds to several seconds.

DISSERTATION OVERVIEW

We start to investigate NC chaperone activity in rearranging nucleic acid secondary structures in the minus-strand transfer step of HIV-1 reverse transcription. This step is of great importance to the whole process of viral replication cycle. We studied the annealing reaction of TAR DNA/TAR RNA (and cTAR DNA) including the reaction mechanism as well as the chemical kinetics using SM-FRET approach. To eliminate the possible large scale aggregates formed by multiple copies of NC/nucleic acids, we applied a novel flow system as a reaction chamber to probe the reaction in an aggregation free manner, which allow us to unravel new insights of the annealing mechanism in the presence of NC. The data demonstrates the partially opened "Y" form of TAR hairpins are important reaction intermediates. Both of the two encounter complexes which are melted TAR hairpins associated with multiple NC annealed to form fully complementary duplex as the final product. The details are given in Chapter 2.

The observation of TAR DNA and its complementary cTAR DNA annealing reaction rate depending on NC concentration promotes us to further explore the reaction mechanism. We first examine the NC concentration dependence on melting of the DNA hairpins. The melting curves for various TAR hairpins were obtained, and the shape of the curve is strongly correlated to the secondary structure of the hairpins. Next, we perform a series of TAR DNA/cTAR DNA annealing reaction under a wide range of NC concentration. The observed rate constant as a function of NC concentration strongly correlated to the melting curve of TAR DNA/cTAR DNA. We also mapped the

thermodynamic stability of different regions on the TAR hairpins. The results demonstrated that the region with less thermodynamic stability melts first and those regions are the "hot spot" for the initial annealing of two complementary DNA hairpins. The details are presented in Chapter 3.

NC is one of the fifteen viral proteins which plays an essential role for HIV. NC is unique because it is a structural protein for the virus architecture and on the other hand it catalyzes multiple steps of nucleic structural rearrangement reaction. We compare and contrast NC which is from zinc-finger binding motif family to the other important family of nucleic-acid-binding protein in HIV, the arginine-rich-binding-motif (ARM) protein, Tat in this case. We employed TAR DNA+cTAR and TAR DNA+TAR RNA annealing assays as a approach to investigate how different NC and Tat's effects are on the nucleic acid hairpin structures. The results have shown that NC can weaken the nucleic acids basepairs while Tat strengthens the stem-loop structure of its target RNA. In the presence of both Tat and NC, annealing of TAR+ TAR RNA can be effectively inhibited. The results may shed light on understanding of the regulation mechanism of the viral gene expression. Detailed data and discussion are presented in Chapter 4.

In the last Chapter, we try to answer a basic question, what is the NC effect on single-stranded DNA (ssDNA)? Based on previous studies, NC is proven to have binding preference to ssDNA, when NC binds onto double-stranded DNA, it tries to melt the weak regions and create more single-stranded regions. Therefore understanding of NC effect on ssDNA conformational change is important for us to explore NC's role on interaction and protection of possible DNA reaction intermediates and genomic RNA molecules. The results show that NC can effectively reduce the rigidity of ssDNA and the interaction between NC and ssDNA molecules are very complex. The results and discussion are given in Chapter 5.

Chapter 2 Detailed study of NC chaperoned annealing reaction mechanism in HIV-1 reverse transcription

INTRODUCTION

During HIV-1 reverse transcription, an essential step in retroviral replication, single-stranded viral genomic RNA is reverse transcribed into double-stranded DNA. Reverse transcription of the IV-1 genome RNA requires two strand transfer steps. During the first strand transfer reaction, the newly synthesized minus-strand strong-stop DNA [(-)SSDNA] is transferred to the homologous sequence in the 3'-untranslated region of the RNA genome, by hybridization of complementary sequences located at the 3' end of the (-)SSDNA and genomic template, respectively (14, 44, 45). In HIV-1, the major components of (-)SSDNA transfer are the stem-loop structured TAR RNA element, its complementary TAR DNA element and the nucleocapsid protein (NC) NCp7, which chaperones this annealing process and other analogous processes in the viral reverse transcription (2, 25, 46-48).

The chaperone (catalytic) activity of NC is derived from two main consequences of the nucleic acid/protein (NC) interactions in this system. First, NC lowers the barrier for annealing by partially melting the basepairs of the hairpins (12, 22, 24, 25, 49). Second, NC lowers the energy cost of bringing the hairpins together in the encounter complex by screening the negative charges of the DNA backbones and perhaps through other specific interactions (16, 23, 27, 29, 30). It has been extremely challenging to develop a clear understanding of the mechanistic role of the NC/nucleic acid interactions because of the tremendously diverse sets of nucleic-acid/NC complexes that have been observed for these systems in vitro. These challenges have been particularly well documented for the NC annealing of the TAR RNA of the HIV-1 genome to the

complementary sequence TAR DNA in the (-)SSDNA (4). Single-molecule spectroscopy (SMS) is uniquely capable of unraveling molecular structure, dynamics, and kinetics in highly heterogeneous and complex biological systems (7, 10, 31, 50). And Single-molecule fluorescence resonance energy transfer (SM-FRET) approaches have been applied on studies of single TAR DNA hairpins (in the absence of TAR RNA or complementary cTAR) structure dynamics in the absence/presence of NC. Previous results have shown that NC induces a shift of the secondary structure of TAR from a fully closed conformation to a partially open conformation, in which the L1L2 stems regions are melted while the L3L4 stems are closed (12, 49). Liu *et al* (16) proposed annealing nucleation can occur through both the zipper and loop mechanism. In their study, the used short oligonucletides to TAR to mimic the initial annealing step at the 3'/5'ternimis and the hairpin loops in the L3L4 region. Intermediates associated with both mechanisms were observed in the presence of NC, and the kinetics of formation of these intermediates was also measured.

In this chapter, we use SM-FRET kinetic measurements to investigate the mechanism of the annealing kinetics of immobilized TAR hairpins to full-length complementary non-immobilized cTAR DNA in the presence of NC. We directly observe a diverse set of annealing intermediates including single NC-coated hairpins, NC-bound complexes of pairs of DNA/RNA hairpins, and large-scale NC/nucleic acid aggregates containing many thousand RNA, DNA, and NC molecules. This tremendous structural diversity is shown to make these reactions unsuitable for analysis by the usual ''homogenous'' reaction mechanistic approaches to identify intermediates and transition states for a reaction. By employing a combination of time resolved SM-FRET techniques, and by simultaneously controlling large scale aggregation, we have been able to address several outstanding mechanistic issues for minus-strand transfer, including the degree of

complexation and secondary structure of the reactants in situ and the identity of the site of the nucleation of annealing (19, 50-54). Then we extend the study to investigate more complicated situations in which single-molecule kinetic data are acquired while the immobilized TAR DNA hairpins are exposed to a time-programmed concentration sequence of different targeted oligonucleotides with and without NC present in the solution. This procedure chemically ‘‘drags’’ individual pairs of reacting hairpins through the reactant states and the intermediate states, and back again (20). This approach offers information on the different stages of the annealing mechanism, especially the putative nucleation complexes.

MATERIALS AND METHODS

Sample Preparation

HIV-1 NC was synthesized as described (16). Functionalized DNA hairpins (purchased from TriLink Biotechnologies, San Diego, CA) and RNA hairpins (purchased from Dharmacon RNA Technologies, Lafayette, CO) were used without further purification as described (Fig. 2.1). Three syringe pumps delivered three solutions, NC, nonimmobilized hairpins (containing Mg^{2+}), and buffer A (containing Mg^{2+} as well), separately. The rapid mixing of solutions was achieved by injecting solutions at 10 μL/min for at least 10 minutes, and then reducing the flow rate to 1 μL/min for the remaining reaction. Acquisition of images and the solution injection were started at the same time. All of the solutions contained buffer A (40 mM NaCl, 25 mM Hepes, pH 7.3 and glucose oxygen scavenger system) (1% v/v 2-mercaptoethanol (Sigma-Aldrich, St. Louis, MO), 3% w/v ß-D(+)glucose (Sigma-Aldrich), 0.1 mg/mL glucose oxidase (Roche

Applied Science, Hague Road, IN), and 0.02 mg/mL catalase (Roche Applied Science) (16).

Coverslip and Chamber Preparation

Coverslips were cleaned in piranha solutions (25% H_2O_2 and 75% conc. H_2SO_4) for 1 hour, followed by various water (molecular biology grade), and acetone (HPLC grade) rinsing cycles. Dry, clean coverslips were then treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for 5 minutes. Coverslips were then rinsed with H_2O and dried under a N2 stream. The clean coverlips were masked with patterned silicone films. The unprotected area was incubated with a 33% w/w polyethylene glycol solution (MW 2000, Nektar Therapeutics, Huntsville, AL) containing 0.25% w/w biotinylated polyethylene glycol (MW 5000, Nektar Therapeutics, Huntsville, AL) in a 0.1 M sodium bicarbonate solution (HyClone, Logan, UT) for 3h. The silicone templates were removed, the excess PEG rinsed with water, and the cover slips dried under a N2 stream. Predrilled polycarbonate films with an adhesive gasket (Grace Bio-Labs, Bend, OR) were assembled on top of cleaned coverslips yielding a chamber with a total volume of ~5 μL. The chamber was assembled on top of the PEG treated surface; the adhesive gasket adhering to the silicone template protected regions of the PEG treated coverslips. Inlet and outlet ports (NanoportTM, Upchurch Scientific, Oak Harbor, WA) were glued on top of the chambers.

Experimental Methods

SM-FRET data were recorded at room temperature by a home-built sample scanning confocal microscope with separate detection channels for detecting Cy3 and Cy5 emission (Fig. 2.2). The Cy3 and Cy5 fluorescence intensity was synchronously detected while rapidly switching the laser excitation between 514 nm and 633 nm, which

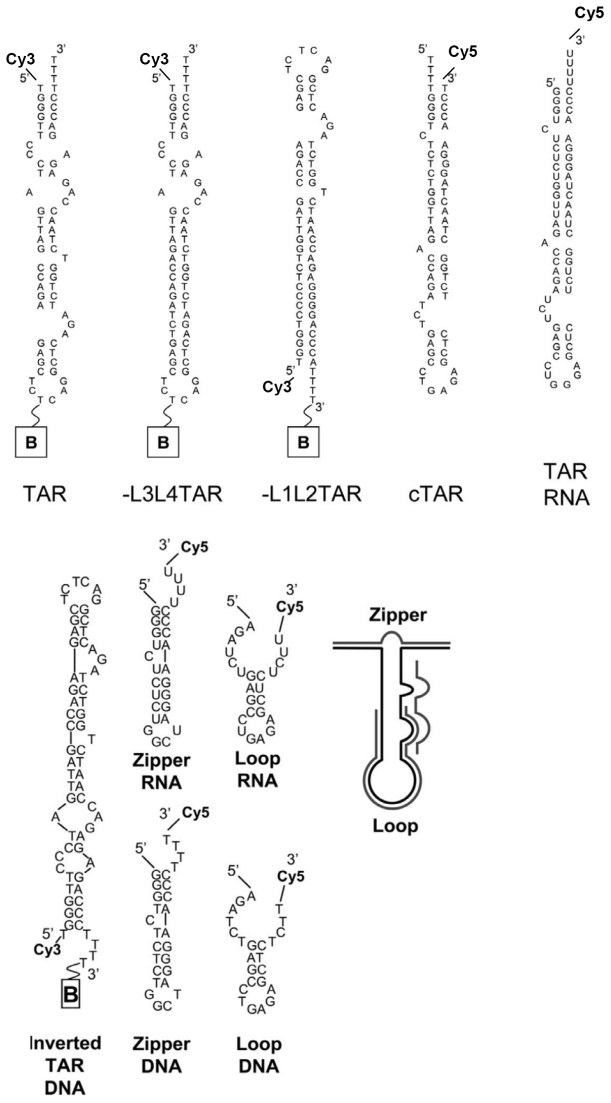


Fig. 2.1 Structures of various oligonucleotides used in this study. The secondary structures were predicted by the program mfold (www.bioinfo.rpi.edu/applications/mfold/dna/form1.cgi) (21).

selectively excited Cy3 and Cy5, respectively. Spatial analysis of the confocal images allowed for colocalization of the Cy3- and Cy5-labeled hairpins. This procedure also yielded the intensity of the ‘‘free’’ Cy5-TAR in the flowing solution above the

immobilized hairpins (which is detected as a 633 nm excited constant background in the images). By monitoring the Cy5 intensity resulting from direct excitation at 633 nm, it was possible to measure the number of Cy5-labeled hairpins, N_{cTAR}, associated (but not necessarily annealed) with a specific immobilized Cy3-TAR at various times during the annealing reaction.

RESULTS AND DISCUSSION

Annealing of TAR to full-length complementary DNA hairpins

Annealing reaction was investigated between nearly all of the pair wise combinations of immobilized and non-immobilized hairpins that are shown in Fig. 2.1. The various hairpins investigated include various mutants that were designed to probe how the kinetics depends on the secondary structure of the DNA hairpins. Unless otherwise noted the annealing experiments employed NC concentrations of 890 nM and nucleotide/NC molar concentration ratios < 5 ensuring saturation binding of NC to the hairpins (51, 55, 56).

Immobilization of the TAR hairpins on a biologically compatible coverslip located in a flow chamber with biotin-streptavidin approach was employed (16) in order to keep one of the annealing reactants stationary and thereby allow for SM-FRET measurements at specific brief time intervals over the duration of the reaction, i.e. typically hundreds to thousands of seconds. The experimental design used a multiple syringe pump flow system allowing for rapid mixing of the NC with non-immobilized hairpin and then rapid delivery of this solution to the reaction chamber (see Materials and Methods). This procedure effectively suppressed formation of large NC/nucleic acid aggregates (in most cases). Aggregation has previously reported to be a serious obstacle

to making direct in vitro measurements of the annealing kinetics for wild-type NC that contain the N-terminal domain, so-called aggregating domain (19, 57). Various DNA/DNA annealing were investigated as a mechanistic comparison and to take advantage of the easier access to DNA mutants and the smaller tendency of DNA to aggregate with NC compared to RNA (58-60).

The immobilized hairpins were labeled at the 5' end with a Cy3 dye (FRET donor dye) and the complementary non-immobilized hairpins were labeled at the 3' end with Cy5 dye (FRET acceptor dye), with the usual attachment configuration (16). Cy3/Cy5 is a well-established donor/acceptor dye pair allowing for an instantaneous measurement of the inter-hairpin Cy3-Cy5 distance by SM-FRET (30, 36, 61). Irreversible annealing kinetics were initiated by exposing a dilute immobilized sample of the Cy3-TAR to a "fresh" solution of Cy5-hairpins containing NC at zero time, t = 0, in analogy to a stopped flow experiment, as shown in Fig. 2.2 for the NC catalyzed annealing of TAR to cTAR DNA. The fluorescence spots in the confocal images for the Donor and Acceptor channels were used to determine the Donor and Acceptor intensities, $I_D(t)$ and $I_A(t)$, and the apparent FRET efficiency, E_A, defined as:

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \qquad (1)$$

By measuring EA for each hairpin, at various times, t, after introducing the complementary Cy5-hairpin solution, FRET trajectories (Fig. 2C) were recorded to monitor the instantaneous distance between the 5' end of the immobilized Cy3-TAR hairpin and the 3' end of the Cy5-hairpins. The time spacing between FRET points (i.e. confocal images) was varied during the experiment to minimize photobleaching. Typically, the time spacing was > 120 s, i.e. the time to record one confocal image.

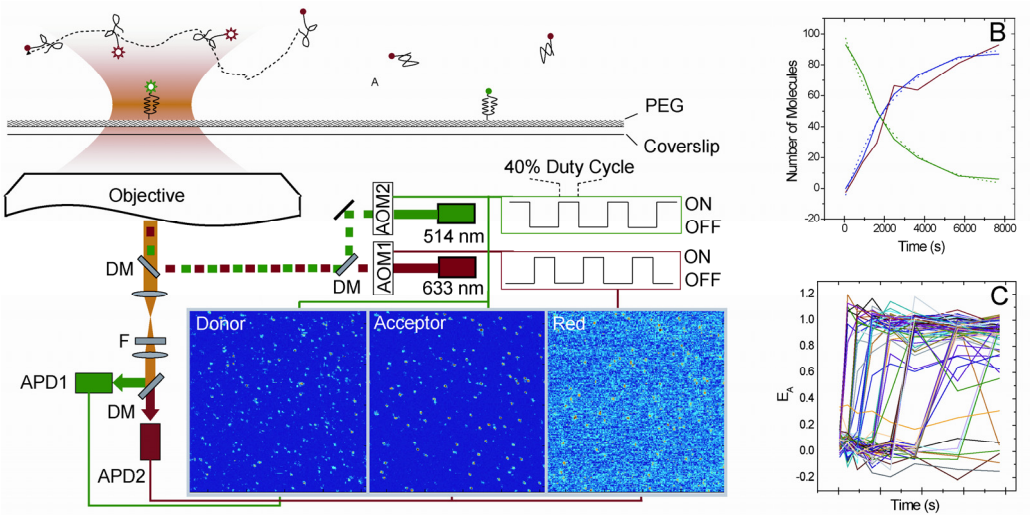


Fig 2.2 A typical single molecule kinetic measurement of TAR and 2 nM cTAR annealing reaction at 0.2 mM Mg^{2+} and 889 nM NC in buffer solution.
(A) The alternating, two color, laser excitation microscopy setup. (B) The number of molecules in three channels (donor, acceptor, and red) during the course of the annealing reaction. (C) EA of each single molecule as a function of time, where each colored line corresponds to a single molecule. AOM, acoustic optical modulator; DM, dichroic mirror; F, notch filter; APD, avalanche photodiode; PEG, poly(ethylene glycol). Modulators controlled by two 180° out-of-phase square-wave signals give two-color alternating-laser excitation. After being filtered by a notch filter, the fluorescence is detected by APD1 and -2. Donor and acceptor channel fluorescence detected by APD1 and -2, respectively, is counted by two counters while the green excitation laser is on. Red channel fluorescence is also detected on APD2, except that it is counted by a third counter while red excitation laser is on.

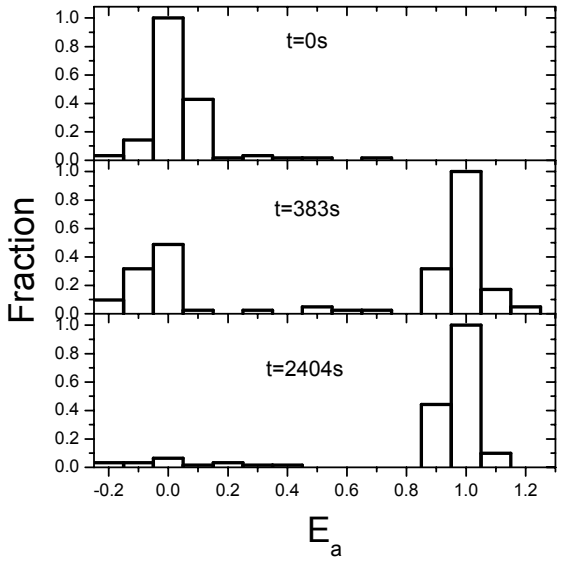


Fig. 2.3 E_A histograms constructed at different time during the annealing reaction of Cy3 TAR with Cy5 cTAR (5nM) in the presence of NC (800nM) ([Mg^{2+}]=0.2mM).

In the presence of both NC and full length complementary cTAR in solution, the observed EA trajectories exhibit discrete jumps from an EA value near zero to a value near unity (Fig. 2.2C) at various times during the annealing reaction. With either or both of NC and complementary hairpins absent the EA value does not vary from zero over the entire time scale of the experiments (data not shown). The discrete annealing trajectories trajectories are especially apparent in "kinetic" histograms of the E_A determined for specific time windows after the initiation of the reaction for an ensemble of reacting hairpins (Fig. 2.3). Thus, the SM-FRET kinetic results directly reveal, in a way that would not be possible with ensemble measurements, that the annealing reaction evolves to kinetic "stable-states", i.e. all other nucleic acid rearrangements must occur on a more rapid time-scale. The mean FRET, $<E_A>$, as a function of reaction time for an ensemble of immobilized hairpins is shown in Fig. 4A. This SM-FRET method for determining

$<E_A>$ is far superior to ordinary bulk solution methods that are subject to artifacts such as dye mislabeling of hairpins, which can easily and automatically be removed during data analysis.

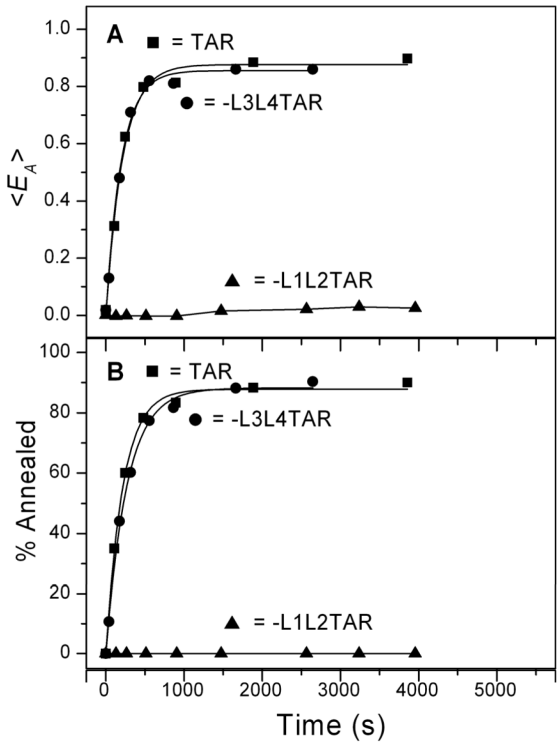


Fig. 2.4 SM-FRET measurements of the annealing kinetics. $<E_A>$ (A) and percent annealed (B) vs. time for the annealing of cTAR (10 nM, non-immobilized) with immobilized wild-type TAR and mutant TAR DNA hairpins in the presence of NC (890nM).

The number of reactant and product molecules, N_R and N_P, respectively were counted directly by using a FRET threshold of 0.4 to distinguish between reactants and products. These N_R and N_P data for TAR/cTAR annealing under saturated NC binding (green and blue solid lines, respectively) were well-fit with a single exponential function (dashed-lines), which is the predicted kinetic behavior for an irreversible bimolecular reaction under the pseudo-first order conditions for these non-immobilized hairpin. The

fitting parameters include a pseudo first order rate constant, k_{ψ}, and an annealing percentage, the latter of which is close to 100%. The single exponential behavior suggests the existence of a kinetic bottleneck for annealing. In chemical kinetics, a single single bottleneck often implies a single well-defined transition state. In the case of the annealing reaction, the conclusion of a well-defined bottleneck applies to the aggregation free (homogenous solution) mode of the annealing reaction. A more complex mechanism mechanism that involves more than one kinetic bottleneck applies when aggregation is not controlled, see below.

A distinct advantage of the confocal method for single molecule spectroscopy is that it allows for synchronous detection of the Cy3 and Cy5 fluorescence intensity while rapidly switching the laser excitation wavelength between 514 nm and 633 nm, which are respectively the wavelengths that excites Cy3 and Cy5. As shown in Fig. 2.2 the 633 nm excitation induces only Cy5 fluorescence (denoted by Red), which is a quantitative measure of the number of bound Cy5 labeled hairpins at various times during the reaction, after calibration. The SMS kinetic data in Fig. 2B shows that the number of associated cTAR hairpins of any type (red curve) is equal to the number of annealed hairpins (blue curve) throughout the annealing reaction, within experimental error. This demonstrates that cTAR is not significantly associated with the TAR until annealing has occurred. The same type of experiments with the other immobilized hairpins mutants in Fig. 1, gave analogous results. An exception to this behavior occurs during large scale NC induced nucleic acid aggregation, see below.

NC/Nucleic acid aggreation

Under conditions in which large scale NC/nucleic acid aggregates are present, e.g. at high cTAR concentration, the observed rate constants for NC induced annealing for the

various hairpins fluctuated greatly from trial to trial. This is reflected in Table 2.1 by the extraordinarily large standard deviations for TAR/cTAR annealing under high cTAR concentration. A number of hypothetical origins for the rate fluctuations such as sticky syringes, temperature variations, and surface absorption of NC or the nucleic acids were ruled out by careful controls. Ultimately, the reaction rate fluctuations were assigned to the aggregation of the cTAR with NC in large molecular aggregates that contain hundreds to thousands of cTAR and NC molecules.

Direct evidence for NC induced large scale of aggregation was observed by recording time resolved fluorescence intensity curves for Cy5-labeled cTAR in solution during annealing reaction. In the absence of NC, the recorded fluorescence intensity only exhibit fluctuation due to photon shot noise and Cy5 cTAR concentration fluctuation. In the presence of NC, intense blips are observed due to single aggregates containing thousands of copies of Cy5 hairpins (17). Aggregation was observed to be especially severe for solutions with high hairpin concentrations, and more severe for RNA than DNA.

It was possible to suppress formation of aggregates in the presence of NC for small nucleic acid concentrations by using high flow rates (freshly mixed solutions). This was further confirmed by fluorescence correlation spectroscopy (FCS) measurement on Cy5-cTAR in the absence and presence of NC. The FCS data in both cases are well-fit by the standard model with the diffusion constants that are consistent with expectations for a single DNA hairpin (17).

The observed huge fluctuations of the annealing rate at high cTAR concentrations were due to the aggregation phenomenon. Over this range the kinetic order of the reaction for cTAR, was observed to vary from first-order at < 20 nM cTAR

concentrations, to zero-order at concentrations in the 25-50 nM range, and then ultimately to exhibit negative orders when the reaction stalled at high cTAR concentration (Table 1).

Table 2.1. The apparent second-order TAR:cTAR annealing rate constants, k_a, at various NC, Mg^{2+}, and cTAR concentrations.[a]

$[Mg^{2+}]$ (mM)	[NC] (nM)	[cTAR] (nM)	k_a ($M^{-1}s^{-1}$)	[NC]/[nt]
2	890	5	$(3.64 \pm 0.34) \times 10^5$	2.8
		10	$(2.72 \pm 1.2) \times 10^5$	1.4
		20	$(3.21 \pm 0.31) \times 10^5$	0.7
1	890	10	$(6.66 \pm 0.47) \times 10^5$	1.4
0.2			$(8.37 \pm 0.048) \times 10^5$	1.4
0.2	500	50	$(1.69 \pm 0.09) \times 10^5$	0.16
		83.3	$(9.16 \pm 10.9) \times 10^2$	0.09
		> 500	No Annealing	< 0.016
0.2	300	50	$(3.73 \pm 2.77) \times 10^4$	0.09
	< 200		No Annealing	< 0.063

[a]The kinetics were measured under pseudo first order conditions, where the [cTAR] >> [TAR]. The *apparent* second-order rate constant, k_a was estimated by dividing the pseudo first order rate constant by [cTAR]. In fact, the reaction is only second-order at high values of [NC]/[nt], see text for further detail. The reaction conditions are: buffer A (40 mM NaCl, 25 mM HEPES, pH 7.3 and glucose oxygen scavenger system) at room temperature. k_a is the average of at least three trials for each reaction.

Annealing Rate Trends in the Absence of Aggregation

Large scale of NC/nucleic acids aggregation can be effectively avoided for low cTAR concentrations (i.e. < 10nM). Under these conditions the data in Tables 2.1, Table 2.2 and Fig.2.4 reveal several mechanistically informative trends. The TAR/cTAR

reactions are observed to be overall second-order, i.e. the pseudo first rate vs. cTAR is linearly proportional to [cTAR]. For TAR/cTAR annealing the data clearly shows that the L1L2 bulges of TAR are required (and in fact are sufficient for rapid annealing). In contrast, removing the L3L4 bulges has little kinetic consequence. While removing the L3L4 bulges does not stop the rapid initiation of annealing, it apparently shifts the annealing equilibrium toward reactants.

Table 2.2 The effect of loop removal on the second-order rate constants, k_a for TAR DNA:cTAR annealing in the presence of 890 nM NC and 2mM Mg^{2+}.[a]

Non-immobilized DNA	**Immobilized DNA**	**k_a ($M^{-1}s^{-1}$) Avg. ± sd. Single exponential fitting[c]**
cTAR	TAR[b]	$(2.7 \pm 1.2) \times 10^5$
	-L3L4TAR	$(3.0 \pm 0.6) \times 10^5$
	-L1L2TAR	No Annealing
-L3L4cTAR	TAR	$(2.2 \pm 1.0) \times 10^5$
	-L3L4TAR	$(2.0 \pm 0.7) \times 10^5$

[a]The reactions were run in buffer A (40mM NaCl, 25 mM HEPES, pH 7.3 and glucose oxygen scavenger system) at room temperature. The *apparent* second-order rate constant, k_a was estimated by dividing the pseudo first order rate constant by the [cTAR]. k_a is the average of at least three trials for each reaction. [b]For TAR:cTAR annealing at different concentrations of cTAR and Mg^{2+}, please see table 1 for the details. [c]For -L3L4TAR:cTAR annealing, a bi-exponential fitting curve can be obtained, which gives a fast rate, $(3.52 \pm 0.75) \times 10^5$ ($M^{-1}s^{-1}$) and a slow rate, $(5.33 \pm 2.71) \times 10^4$ ($M^{-1}s^{-1}$).

By using a flow system with rapid mixing, we have been able to investigate the annealing kinetics with full length NC in an aggregation free condition. Previously, in order to suppress aggregation effects on the annealing kinetics it was necessary to use a truncated form of nucleocapsid protein, NC(12-55), which lacks the so-called aggregating N terminal domain (57)

Based on the new information obtained, a refined and more specific mechanism for the NC chaperoned annealing of TAR to cTAR/TAR RNA has been proposed (Fig. 7). This mechanism applies to the annealing reaction in the absence of large scale NC/nucleic acids aggregates. The SMS results demonstrate that the TAR reactant is predominantly a single NC coated, hairpin with a dynamic secondary structure, involving a partially-open "Y" shaped conformations (equation (i) of Fig. 7). It is reasonable, based on the experiments for TAR, to assume that the cTAR or TAR RNA reactants also have a partially melted secondary structure, which provides the exposed single stranded region for nucleation of annealing.

The mechanism portrayed in Fig. 2.7, assumes that rapid and reversible association and partial melting precedes a much slower nucleation event. The observation herein that the L1L2 bulges of TAR are required (and in fact are sufficient) for rapid annealing, coupled with the previous observation that hairpins missing the L1L2 bulge region do not undergo NC induced melting strongly suggests that the encounter complex is formed by two hairpins with at one or both of the hairpins in the "Y" conformation (12, 13). This conclusion is reflected in the first step in equation (iii). The rate-limiting-step (RLS) for annealing, is hypothesized to be local annealing at specific locations along the hairpins, which can "nucleate" the annealing process. A steady-state solution of this mechanism predicts the second-order overall annealing kinetics (in the absence of large scale aggregates), which is in agreement with the experimental results. Furthermore, the

close similarity of the annealing rate constant of TAR with cTAR, to that for other oligonucleotides that should strongly favor the zipper mechanism (e.g. -L3L4TAR and TAR) strongly suggests that the zipper mechanism dominates the annealing reaction.

We have previously shown that a short, locally targeted oligonucleotide for the hairpin loop, including the L3L4 regions, is capable of annealing, although at a slower rate than the zipper nucleation route (16). This implies that there may be two pathways for annealing for full length cTAR and TAR RNA. However, for the full length oligonucleotides, nucleation by either route ultimately leads to the same duplex annealed product. The similar rates of the loop and zipper routes (only an order of magnitude slower at low Mg^{2+} concentrations) suggests that a common mechanism may be operating in all cases. Nucleation of annealing at two different locations corresponds to two distinct transition states for the reaction. It should be emphasized, however, that the two pathways are highly analogous, involving nucleation of annealing in locally single-stranded regions in the NC melted and associated complex of the two hairpins. Thus, a better description for the reaction than two-transition states may be a reaction mechanism in which the transition state is a broad multidimensional region along a "wide" reaction path with two or more, adjacent forms of the same type of transition states corresponding to different nucleation sites. Such complexities are not surprising in nucleic acid arrangements due to the rugged energy landscape (47, 62, 63).

The SMS results strongly suggest that the common approach of investigating the NC induced annealing kinetics of nucleic acids with aggregates present must be undertaken with great caution due to the exceedingly complex and heterogeneous nature of these reactions at the molecular level. Nevertheless, the annealing reaction kinetics and mechanism within large scale aggregates is of interest due to its potential similarity with certain aspects of the annealing process in vivo.

Reversible annealing of TAR to oligonucleotides in the presence of NC

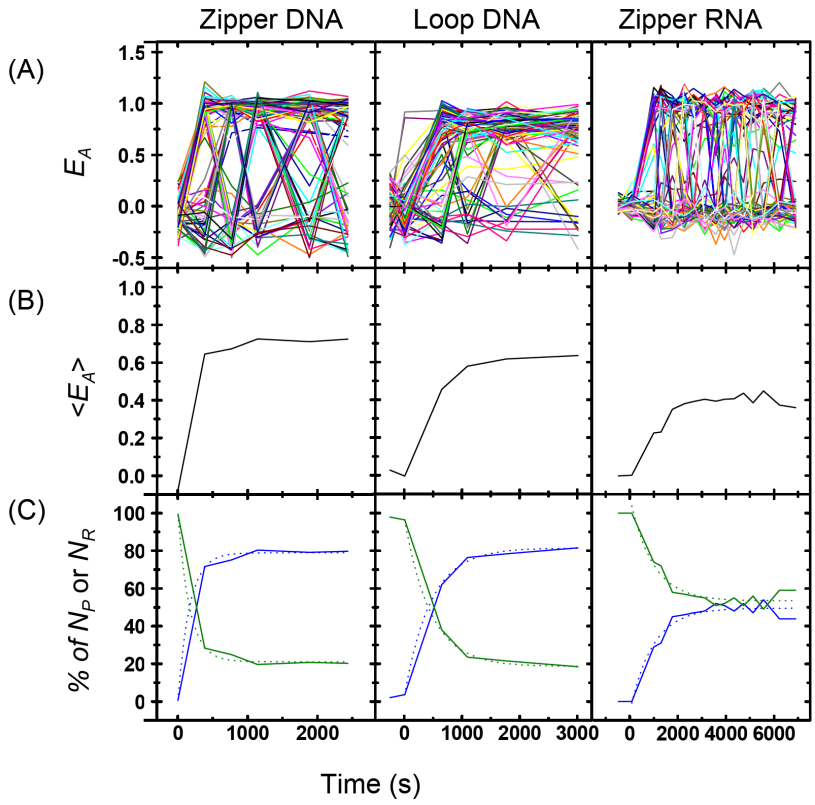


Fig. 2.5 SM-FRET kinetic measurement on the annealing of Cy3-TAR DNA to 25 nM Cy5-zipper DNA, 25 nM loop DNA and 50 nM Cy5-zipper RNA at 0.2 mM Mg^{2+} and 889 nM NC. Top panels show the EA of each single molecule during the reaction, where each colored line corresponds to a single molecule. Middle panels show the ensemble mean E_A increases during the annealing reaction. Bottom panels show the number of product (N_P, shown as % among total number of molecules) molecules during the annealing reaction (blue, solid), while the number of reactant (green solid) (N_R) molecules decrease by the same amount. Both of the increase/decrease trends over time can be fitted with a single exponential (dotted lines). The data was collected from 100 single molecules for cTAR annealing kinetic measurement, 140 for zipper annealing and 200 for loop annealing.

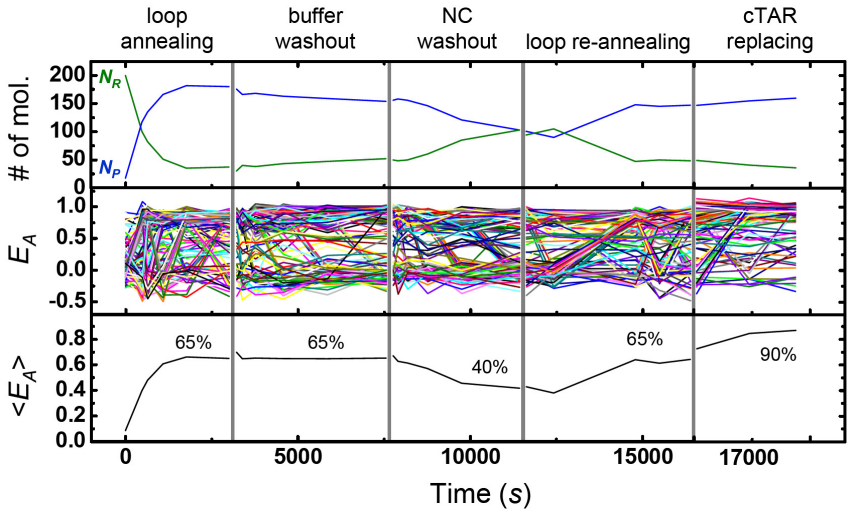


Fig. 2.6 SM-FRET kinetic measurement of the annealing reaction between immobilized Cy3-TAR DNA and 25 nM Cy5-loop DNA in the presence of NC (889 nM). The sample was then washed out by buffer and NC sequentially, re-annealed with 25 nM Cy5-labeled loop DNA (re-annealing) and reacted with 25 nM Cy5-cTAR DNA (cTAR). All solutions contained 0.2 mM Mg^{2+}. Top panels show the number of annealed (blue) and un-annealed TAR DNA (green) molecules at different reaction stages. Middle panels show the single molecule EA trajectories, where each colored line corresponds to a single molecule. Bottom panels show the mean E_A of the single molecules in the middle panel. The final annealed percentages in each reaction stage are also shown in the corresponding graphs.

After investigation of the irreversible annealing of TAR to full length cTAR DNA, we focused on the multi-component oligonucleotides SM-FRET experiments in this section. Two types of short oligonucleotides which are designed to target the "zipper" and "loop" intermediates respectively were employed in this study. The typical annealing reaction kinetic curves are shown in Fig. 2.5. The bottom row (Fig. 2.5) portrays the number of surviving reactant hairpins N_R (green curve) and the number of annealed product N_P (blue curve) pairs. These were determined by counting the number of hairpins that were below/above an E_A threshold of 0.4, as described above.

The second and third column in Fig. 2.5 correspond to annealing experiments experiments with short DNA oligonucleotides that are targeted for "zipper" (L1L2 stem

loops) and "loop" (L3, HL) regions of TAR DNA. SM-FRET data on annealing of both target oligonucleotides show evidence of reversible annealing that leads at later times to an equilibrium distribution of annealed and unannealed TAR. For example the individual individual E_A trajectories show much more high-to-low FRET transitions than the TAR DNA/cTAR case. The $<E_A>$ curves and the number of reactant and product data are also also consistent with an equilibrium mixture at long times, with an apparent dissociation constant K_d of ~ 10 nM and ~ 16 nM for the zipper and loop DNA, respectively. One can can estimate K_d from the ~ 70% and ~ 60% annealing percentage at 25 nM concentration for zipper and loop DNA, respectively. The annealed adducts of TAR DNA with the target oligonucleotides are arguably models for intermediates in the annealing reaction of full length cTAR or TAR RNA and indeed, in minus strand transfer itself. The annealing of TAR DNA with TAR RNA, zipper RNA and loop RNA are more difficult to study accurately due to their higher tendency of forming aggregates and sticking onto surface, but generally exhibit similar kinetic behavior to their DNA analogs. As an example, the column of Fig. 2.5 shows the kinetics of TAR DNA/zipper RNA annealing.

In order to characterize the nucleic acid rearrangement pathways available to the TAR DNA/target-oligonucleotides we subjected these models for nucleation-complexes to a sequence of solutions containing buffer only, buffer plus NC, and finally, a solution of NC plus cTAR. Typical results are shown in Fig. 2.6 for the loop DNA oligonucleotides. In the first epoch of the experiment, immobilized TAR was exposed to a loop DNA plus NC solution, leading to an annealing equilibrium. This was followed by a period in which the loop DNA plus buffer solution was rapidly replaced with a buffer only solution. The major effect of the buffer-only period was to freeze the concentration of the TAR/loop DNA adduct, even though the equilibrium constant strongly favors dislocated adducts in the absence of oligonucleotides in solution. In

contrast, when NC was added to the solution (in the third epoch) the concentration of the adduct decreased relatively rapidly and continuously. The final epoch in Fig. 2.6 involves a NC-promoted strand displacement of loop DNA by cTAR. This step in the sequence is a simple assay on the "activity" of the immobilized TAR hairpins, allowing for a validation of the entire procedure. Since efficient cTAR annealing was observed, it is confirmed that the TAR DNA hairpins were not damaged nor poorly immobilized by the programmed sequence of reagents.

This data demonstrates that NC not only catalyzes the forward annealing process, but also catalyzes the reverse annealing process. This is expected according to microscopic reversibility for a NC chaperoned process. Since NC interacts more strongly with single stranded DNA than with double stranded DNA, the combined NC effects strongly suggest that the transition state for the annealing reaction possesses more single stranded regions than the reactants, products, and even the stable intermediates along the reaction path. This is qualitatively consistent with the proposed "Z" nucleation-complex structure in Fig. 2.7.

Various observables and derived kinetic parameters for the annealing reaction of TAR DNA with the various oligonucleotides are listed in Table 2.3. Some clear trends are apparent in the data. For example, NC is clearly required for both the annealing and reverse annealing reactions to be rapid. Also, the reverse annealing of the full length oligonucleotides (cTAR and TAR RNA) are much slower than the short oligonucleotides consistent with the idea that the annealed adduct of the TAR DNA with the short oligonucleotides is indeed a model for the nucleation complex for the full length annealing reactions. In other words, considerably more base pairs must be broken to achieve the proposed transition state for reverse annealing of the full length cTAR.

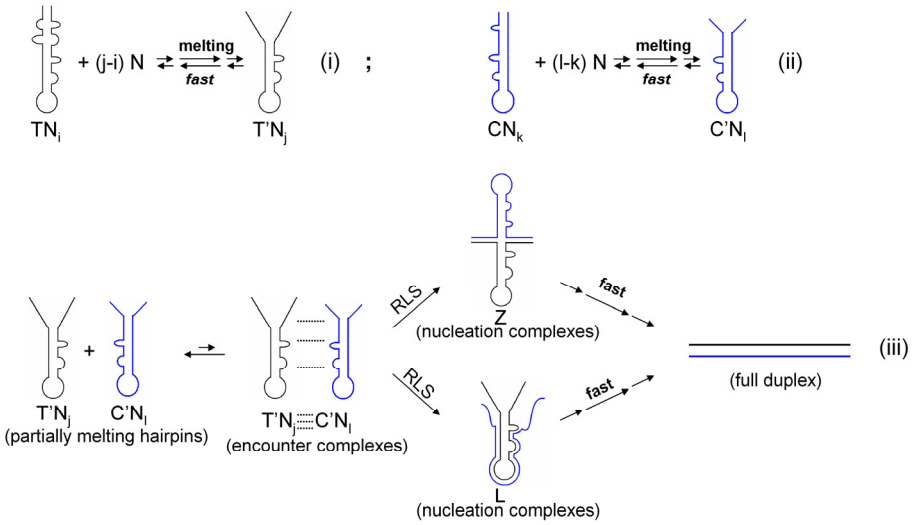


Fig. 2.7 A hypothetic kinetic scheme for NC chaperoned annealing of Cy3-TAR DNA to its Cy5-labeled complements. Here, T denotes TAR DNA, and C denotes complementary cTAR DNA or TAR RNA. The term N denotes NC. In this scheme, N binds to T and C, leading to a partially melted structure, namely the Y form of T (T') and C(C'). The subscripts, i, j, k and l are used to describe the number of NC bound to nucleotides. Two partially melted hairpins form an encounter complex that leads to the formation of nucleation complexes. The annealing can go through either zipper nucleation or loop nucleation, therefore, forming zipper nucleation complexes (Z) or loop nucleation complexes (L) both leading to the formation of fully annealed duplexes.

Further evidence for the reversibility of this pathway was obtained by forming forming an equilibrium mixture of the TAR DNA/zipper DNA adduct by NC annealing and then rapidly replacing the NC plus zipper DNA solution with a NC plus cTAR (Cy5 labeled) solution. The zipper DNA hairpins were observed to be efficiently replaced by the cTAR (data not shown). Presumably, this net "strand displacement" reaction occurred due to NC induced annealing between cTAR and the unannealed TAR DNA. (The unannealed TAR DNA presumably resulted from NC induced reverse annealing of the TAR DNA/zipper-DNA adduct.) Various strand displacement reactions were undertaken as summarized in Table 2.4. In every case where a stable annealed product

between TAR DNA and a full length oligonucleotide was expected, the strand displacement reaction was too slow to observe. For example, it was observed that the cTAR in a TAR DNA/cTAR adduct were not efficiently displaced by a concentrated mixed solution of NC and zipper DNA (and also not washed-out by NC only). Correspondingly, for all adducts of TAR DNA with short oligonucleotides, efficient strand displacements of the short oligonucleotide by another oligonucleotide were observed. All the above observations are under out time resolution (>120s). The faster events beyond the time resolution will be missed and the actual reaction mechanism might be more complex.

Table 2.3. Rate constants for annealing (k_a) and reverse annealing (k_r) at 0.2 mM Mg^{2+} in the presence or absence of 889 nM NC.

Parameter[a]		cTAR DNA	Zipper DNA	Loop DNA[c]	TAR RNA	Zipper RNA	Loop RNA[c]
k_a ($10^5 s^{-1}M^{-1}$)	- NC	0	0	0	0	0	0
	+ NC	8[b]	10	1	2	0.1	0.06
k_r ($10^{-4}s^{-1}$)	- NC	0	1[d]	0.1[d]	0	5	0
	+ NC	0	18	4	0	5	3

[a] k_a denotes the annealing rate constant and k_r denotes the reverse annealing rate constant. The annealing reaction kinetic curve is measured in the presence (+NC) or absence (-NC) of NC and fitted by a single exponential function. The single exponential rate constant k is $k = k_r + k_a \cdot [NA]$ (NA = reacting nucleic acid). k_r is determined by washing out the annealed product with NC (+NC) or buffer only (-NC) solution, and fitting the kinetic curve by a single exponential function with a rate constant k_r. k_a is determined from k and k_r as $k_a = (k - k_r)/[NA]$. [b] value from our previous published data (64). [c] The loop annealing kinetic were measured with an inverted TAR DNA construct. Although there does not appear to be appreciable steric interference for the TAR DNA-loop DNA annealing (65), there is measurable steric interference of the TAR DNA-loop RNA annealing. [d] This value is based on the slow decay component of *Np*. The fast decay component is associated with reverse annealing of 1-armed annealed TAR DNA/zipper or loop DNA adduct.

Table 2.4. Summary of strand-displacement experiments.[a]

REACTION SEQUENCE
------ Efficient Strand Displacement ------
Cy3-TAR/Cy5-Zipper-DNA+ Cy5-cTAR
Cy3-TAR/Cy5-Loop-DNA + Cy5-cTAR
Cy3-TAR/Cy5-Zipper-RNA + Cy5-cTAR
Cy3-TAR/Cy5-Loop-RNA + Cy5-cTAR
------ Inefficient Strand Displacement ------
Cy3-TAR/cTAR + Cy5-cTAR
Cy3-TAR/cTAR + Cy5-Loop-DNA

[a] The annealing and replacement experiments were run in buffer containing 0.2 mM Mg^{2+} and 889 nM NC, 25 nM Cy5-zipper/loop nucleotide and Cy5-cTAR DNA .

CONCLUSION

New insights of NC chaperoned annealing mechanism of a model TAR DNA hairpin sequence to the complementary cTAR (or TAR RNA) hairpin which is an essential step of HIV-1 reverse transcription have been obtained. In combination of SMS fluorescence measurements and a novel flow chamber approach to suppress large scale aggregates, we have explored the secondary structures of key intermediates and transition stats along the annealing reaction pathways. The SMS results demonstrate that the TAR hairpin reactant is predominantly a single NC coated hairpin with a dynamic secondary structure, and the nucleation of annealing occurs in an encounter complex that is formed by two melted hairpins.

In addition, a multi-component SM-FRET kinetic approach has been developed for analyzing the mechanism of NC induced nucleic acid rearrangements that involve multiple oligonucleotides. The approach involves exposing one of the oligonucleotides to a sequence of solutions, containing the complementary oligonucleotide, other target oligonucleotides, buffer-only, and the chaperone protein in

various combinations. This procedure effectively drags the nucleic acid protein systems from reactants through key intermediates, and finally toward the rearranged products, while monitoring the conformational states and dynamics of the system with SM-FRET. The results strongly suggest that the nucleation for annealing of TAR DNA to cTAR DNA involves base pair formation, in different regions of TAR DNA. In addition, the results clearly demonstrate that NC is capable of catalyzing reversible annealing at various stages along the reaction path of the HIV-1 reverse transcription.

Chapter 3 Concentration-dependence of HIV-1 NC chaperoned melting and annealing of DNA hairpins

INTRODUCTION

HIV-1 nucleocapsid protein (NC) is a small, basic viral protein with 55 amino acids, containing two highly conserved zinc fingers with the CCHC zinc-binding motif. (66-69) Playing important roles in almost every step in the viral replication cycle, NC is considered as a multifunctional protein encoded by HIV-1. For instance, NC promotes the dimerization of the genomic RNA and stabilizes the ribonucleoprotein complex, (70-75) involves in the genomic RNA packaging, (76) facilitates the placement of the tRNA primer onto the RNA primer binding site, (77-81) and chaperones (catalyzes) the two obligatory strand transfer events (minus- and plus-strand transfer) during the viral reverse transcription (25, 82-86).

The integrity of the zinc fingers is critical to the NC's function in retrovirus replication, and deletion of one zinc finger or mutation in either zinc finger impair the viral RNA packaging and reverse transcription (87-92). Proper folded zinc-finger structure has been shown to be essential for recognition and interaction of the target nucleic acids, such as stem-loop structured SL3 of packaging signal (93). NMR data have shown that there is specific interaction between the single-stranded loop region and the hydrophobic ring of Trp37 of the zinc-finger. NC is known for its general non-sequence-specific interaction with single-stranded as well as duplexed region of nucleic acids. On the other hand, previous studies also demonstrate NC exhibit binding preference to sequences containing repeated (TG) or (UG) (51, 94).

NC is required and essential for HIV-1 reverse transcription. There are two strand transfer steps involved in the process, in our study; we focus on how NC catalyzes the

minus-strand transfer. The minus-strand transfer step involves the transfer of newly-synthesized minus-strand strong-stop DNA ((−)SSDNA) to the 3′ end of viral RNA by annealing it to the complementary repeat (R) regions of the genomic RNA (44, 45, 95). The major component of the R region is the highly structured transactivation response element named TAR (96, 97), which folds into a stable stem-loop structure. The necessity of nucleic acid chaperone arises because annealing of the TAR to its full length complementary TAR DNA is prevented by the internal self-paring regions of each individual DNA/RNA hairpins. NC chaperons (catalyzes) the annealing reaction of TAR DNA and TAR RNA by rearranging the structures of nucleic acid hairpins therefore lowering the activation energy of forming duplex (23, 98-100). This chaperone activity of NC is believed to arise from two consequences of NC binding to nucleic acids. First, NC lowers the energy barrier for annealing by partially melting the basepairing of the duplex region due to NC preference for binding to single-stranded regions of both RNA and DNA. Second, NC lowers the energy cost of bringing two complementary hairpins together to form encounter complexes by screening off the negative charges of the nucleic acids.

We have previously used SM-FRET to identify melted secondary structure intermediates in the NC chaperoned annealing of TAR DNA to cTAR DNA/TAR RNA (12, 16, 17, 20). The mechanism of TAR annealing reaction shown in scheme 1 has been proposed (see chapter 2). In this study, we extend the melting experiments of TAR (and TAR mutants) and the annealing experiments to a wide range of NC concentration. SM-FRET approach was employed to examine the DNA hairpin secondary structure as well as the TAR DNA + cTAR DNA annealing kinetics as a function of NC concentration.

Widely used gel mobility shift assay were used in this study to analyze the binding properties of NC to TAR DNA and mutants with different secondary structures

(101, 102). Interestingly, we have compared the TAR/NC binding patterns in the absence and in the presence of EDTA in the gel running buffer. EDTA is known as an efficient zinc chelating reagent which destroys the zinc finger structure by ejection of the zinc ion from the protein. However, EDTA was used in previous studies as a component in the gel running buffer TBE. Instead, we used TB buffer (no EDTA) as running buffer in the gel shift assay to investigate NC binding properties of TAR (and mutants) hairpins with different secondary structures (Scheme 3.1). TAR DNA hairpins melting experiments was also performed here to investigate the role of zinc-finger for the NC chaperone activity.

To fully map the thermodynamic stability of different regions within TAR hairpins, we combined the SM-FRET results of previous melting/annealing experiments in the presence of NC. Therefore we are allowed to identify the "hot spot" within the DNA hairpins which are more easily to be melted by NC or annealed with target oligonucleotides in the presence of NC. Additionaly, we used mfold (103) to calculate the free energy of differently folded DNA secondary structures and then compared it to the experimentally determined energetic map. These results help us in understanding the role of NC as a nucleic acid chaperone and its effect on nucleic acid secondary structures at the molecular level.

MATERIALS AND METHODS

Nucleic Acids and Protein Preparation

HIV-1 NC was synthesized as described (16). Functionalized DNA and RNA sequences (Scheme 3.1) were purchased from TriLink Biotechnologies (San Diego, CA) and Dharmacon RNA Technologies (Lafayette, CO), respectively and were purified by

the supplier by polyacrylamide gel electrophoresis and reversed-phase HPLC. Cy3 and Cy5 were used as a FRET pair and were attached to nucleic acids at the 5′ and 3′ ends, respectively. Biotin was also attached in hairpin loop region of each nucleic acid for immobilization to biologically compatible coversilps (immobilization see below).

Coverslip Preparation and Nucleic Acid Immobilization

Clean coverslips were treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for five minutes and then pegylated and biotinylated. A reaction chamber with an inlet and an outlet port (Nanoport, Upchurch Scientific, Oak Harbor, WA) was glued on top of treated coverslip.(12, 13, 16, 104-106) All the nucleic acids were diluted to 50-100 pM, and reannealed by incubation for 2.5 min at 80°C, 2.5 min at 60°C, and 5 min at 0°C before immobilization to coverslips. The buffer solutions, NC and other reactants were delivered into the reaction chambers by syringe pumps with programmable flowing rates. All the solutions contained 25 mM HEPES buffer, pH7.3, 40mM NaCl, $MgCl_2$ (Mg^{2+} concentration specified in different experiments) and an oxygen scavenger system (2-mercaptoethanol 1% v/v, glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL) and catalase 0.02 mg/mL (Roche Applied Science, Hague Road, IN) (40, 107).

Single Molecule Spectroscopy Data Collection and Analysis

The home-built sample scanning confocal microscope used for SM-FRET measurements has been described extensively elsewhere.(12, 13, 16, 104-106) Donor and acceptor images were recorded at the equilibrium conditions for NC concentration titration measurements or in time-resolved mode for annealing kinetics measurements. The donor and acceptor emission intensities ($I_D(t)$ and $I_A(t)$) were taken from the donor and acceptor intensities of each immobilized nucleic acids and corrected for background

emissions and donor/acceptor cross talks. The apparent FRET efficiency, $E_A(t)$, is calculated as

$$E_A(t) = \frac{I_A(t)}{I_A(t) + I_D(t)} \quad (1)$$

The actual FRET efficiency, $E_{FRET}(t)$, of the system is given by

$$E_{FRET}(t) = \frac{I_A(t)}{I_A(t) + I_D(t)\frac{\phi_A \eta_A}{\phi_D \eta_D}} \quad (2)$$

Where ϕ_D, ϕ_A are the donor and acceptor dye quantum yields and η_D, η_A are the donor and acceptor detector efficiencies, respectively. For our experimental setup, the correction factor $\phi_A\eta_A/\phi_D\eta_D$ is ~1, therefore the apparent FRET efficiency $E_A(t)$ is equal to $E_{FRET}(t)$.

Gel mobility-shift Assay

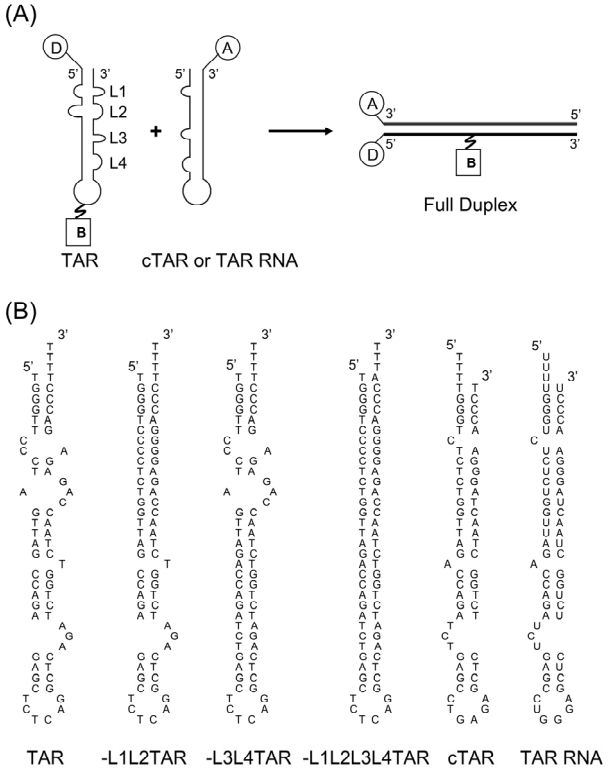
TAR DNA, –L1L2 TAR, –L3L4 TAR, –L1L2L3L4 TAR, cTAR and TAR RNA were labeled with Cy3 and Cy5 at the 5′ and 3′ ends, respectively, and –L1L2 TAR was only labeled with Cy3 at the 5′ end. Binding reaction mixtures of total 20 μl contained 25nM nucleic acids with NC at various concentrations, and were incubated in 25mM HEPES buffer, pH7.3, 40mM NaCl, 0.2 mM MgCl2 for 15 minutes before loading. The samples were run with 10% native polyacrylamide gel at room temperature under 230 V for 2 hours with running buffer of TB (Tris-Borate, 45mM) or TBE (Tris-Borate, 45mM and EDTA, 1mM). Fluorescence images were then recorded with a Typhoon Trio Imager (Amersham Biosciences, Piscataway, NJ) with excitation at 532 nm and emission > 670 nm for FRET mode and excitation at 532 nm and emission > 580 nm for Cy3 mode. Gel images were analyzed with Quantity One (Bio-Rad Laboratories, Hercules, CA) and fitted with Hill equation.

$$Bound\% = \theta_{\max} \frac{[NC]^n}{K_d^n + [NC]^n} \quad (3)$$

Where, θmax is maximum bound percentage, [NC] is NC concentration, K_d is the dissociation constant, n is Hill coefficient.

RESULTS AND DISCUSSION

Melting of DNA hairpins in the presence of NC



Scheme 3.1 Illustration of the annealing reaction of TAR/cTAR or TAR/TAR RNA and various oligonucleotides used the experiments. (A) Schematic representation of annealing reaction of TAR and cTAR or TAR RNA in Single Molecule FRET studies. Cy3 and Cy5 are used for fluorescence donor and acceptor, respectively. TAR DNA has four bulges or loops (L1, L2, L3 and L4), and is immobilized via a biotin linker to a coverslip, cTAR or TAR RNA are flown in with NC to the reaction chamber. (B) Structures of nucleic acids used in the Single Molecule FRET studies and gel mobility-shift assays. All the nucleic

acids are doubly-labeled with the FRET pair Cy3/Cy5 at the ends, except for –L1L2 TAR, which is only Cy3-labeled at the 5' end.

NC-induced melting of different secondary structures, which is achieved by designing a series of TAR DNA mutants, is measured by SM-FRET. Dual labeled oligonucleotides with donor (Cy3) and acceptor (Cy5) dyes at the two ends of the hairpin are immobilized on the biologically compatible surface and exposed to buffer and protein solutions with various NC concentrations. For each NC condition, FRET images are recorded when the systems reach equilibrium. And FRET histograms were constructed from FRET values of more than 300 single molecules. Fig. 3.1 shows the ensemble mean FRET value of these molecules as a function of NC concentration. Apparent mean FRET value, $<E_A>$, reflects the separation between two dyes therefore the information of oligonucleotide melting induced by NC can be extracted. For a closed hairpin, mean FRET is ~1, and for a partially melted "Y" form, mean FRET decreases to a value smaller than 1 depending on the structure of final product. Two main parameters from these melting curves are derived for a specific TAR DNA hairpin. First, the minimum $<E_A>$ value is obtained from analyzing the quasi plateau region of the graph, where increasing NC concentration can no longer melt oligonucleotide (NC saturating conditions). The second parameter is the critical NC melting concentration, C_m, which is simply taken as the NC concentration where half-way mean FRET change occurs.

NC-induced nucleic acid melting highly depends on the secondary structure of the oligonucleotide. The minimum $<E_A>$ value is 0.80 for TAR DNA, 0.78 for cTAR, 0.82 for –L3L4 TAR and 0.99 for –L1L2L3L4 TAR (Fig. 3.1). This data clearly demonstrates demonstrates that NC is capable of rearranging secondary structures of DNA hairpins in a NC concentration dependent manner. There is a critical NC concentration which is

requited to partially open the hairpins under our experimental condition. The degree of melting highly depends on the secondary structure of the DNA hairpins. The number of single-stranded region and their location has a strong effect on its structure rearrangement induced by NC. For a perfectly base-paired hairpin (–L1L2L3L4 TAR), NC's melting ability is rather limited, high concentration of NC (800 nM) still has minimal effect on the hairpin. It is also interesting to note that except for –L1L2L3L4 TAR all DNA hairpins have a Cm of 70~100 nM, and similar NC saturating concentration of > 200 nM.

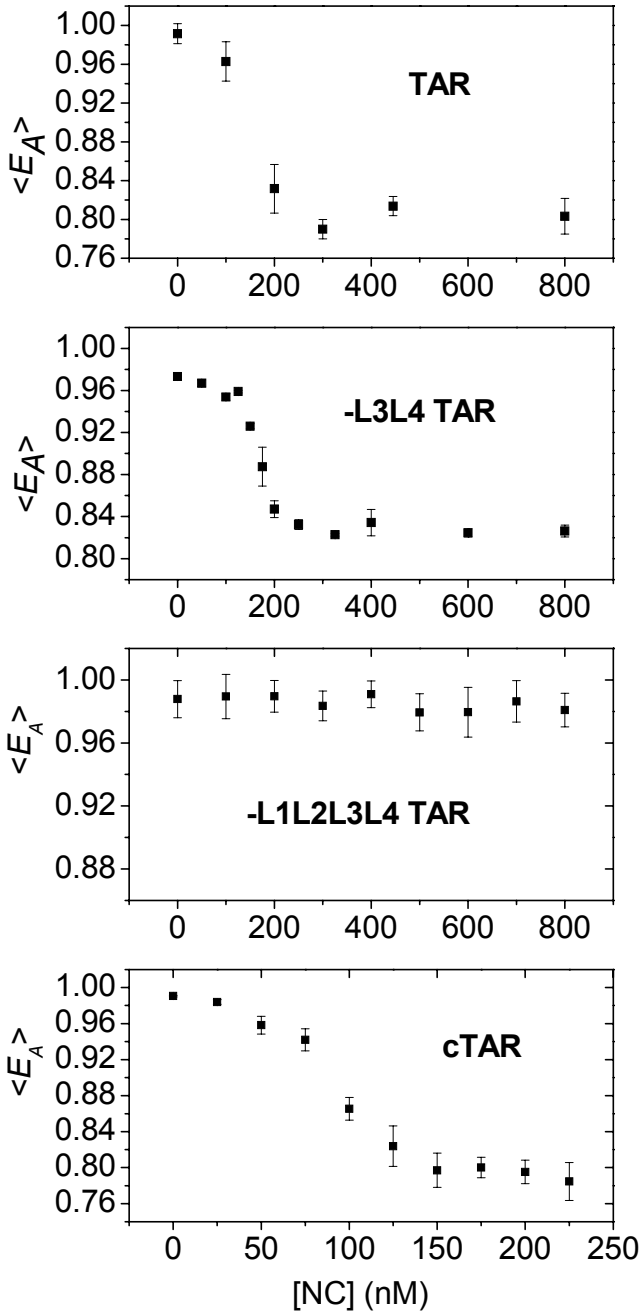


Fig.3.1. NC melting curves of various TAR hairpins. ([Mg^{2+}]= 0.2 mM in solution). The mean FRET values at different NC concentrations are shown for Cy3/Cy5 doubly-labeled (A) TAR, (B) –L3L4 TAR, (C) –L1L2L3L4 TAR, and (D) cTAR.

A dramatic incensement in the TAR +cTAR annealing reaction rate was observed at NC concentration ~100nM, which falls in the same range of Cm of NC induced hairpin melting (Fig. 3.2). The SMS kinetic annealing assay is a more sensitive approach to monitor NC-induced melting, in which the yield of NC catalyzed annealing of TAR DNA with cTAR is recorded as a function of time for a range of NC concentration (0~800nM). Fig. 3.2 shows the observed second-order rate constant as a function of NC concentration. Below 100 nM of NC, there is no apparent annealing product observed under our experimental condition, starting from 100 nM, the annealing rate increases dramatically and starts to reach a plateau after 200 nM. The annealing occurs only after NC melts the reactant hairpins from the closed form to partially open "Y" form; and the annealing assay is an indirect means to monitor the NC-induced TAR hairpins melting process. The critical NC concentration for accelerating TAR +cTAR annealing from SM-FRET kinetics measurements exhibit strong correlation with C_m obtained from NC titration for melting experiments (Fig. 1). This data solidifies our previously proposed kinetic scheme of TAR annealing to its complementary DNA or RNA (Chapter 2). (104, 105) Only melted hairpin structures induced by NC (for both reactants) can lead to the encounter complexes where single-stranded region effectively base-pair to each other, subsequently form the annealed duplex. The melting process induced by NC is required and necessary for the effective annealing to occur, and the annealing rate depends on the degree of the melting of both reactants by NC. For a perfectly double-stranded hairpin, on which NC has minimal melting effect (within our experimental concentration range), annealing does not occur within our observation time window (up to several hours) (104).

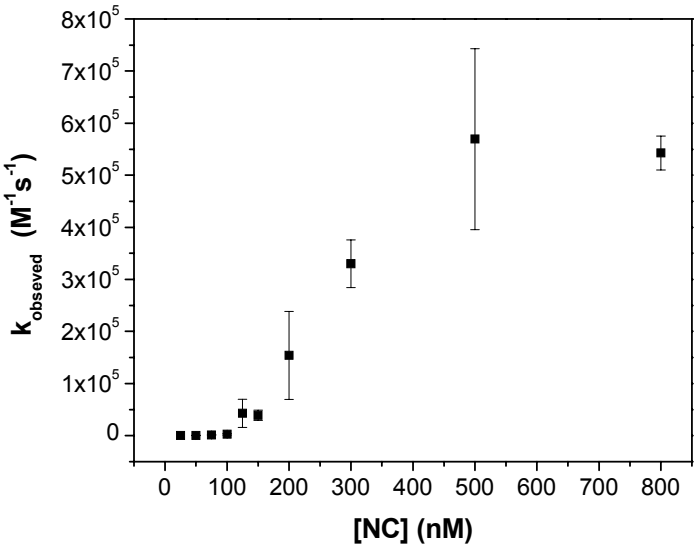


Fig. 3.2. Observed second-order rate constants derived from the SM-FRET kinetics of the annealing of TAR and cTAR at different NC concentrations. Annealing kinetics was recorded at each NC concentration, and was fitted with single-exponential function. The observed second-order rate constant was estimated by dividing the pseudo-first-order rate constant by [cTAR]. Multiple kinetic runs were conducted and error bar was indicated in the plot.

SM-FRET titration experiments show that zinc-finger-depleted NC, where the wild-type NC is treated with 1mM EDTA for at least 30 min, loses its melting ability (Fig.3.3). Data of –L3L4 TAR DNA shows no change by EDTA-treated NC in the concentration range of 0~1000 nM. In addition, NC (11-55) (no N terminal domain) shows limited melting effect on –L3L4 TAR DNA with a much bigger Cm of ~ 600 nM (data not shown). Taken together, these results suggest that full chaperone activity of NC requires both zinc-finger structure and the N-terminal domain.

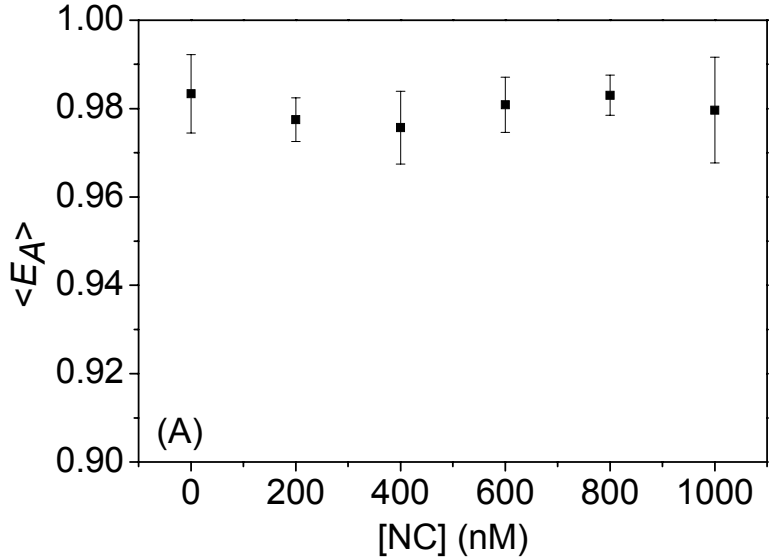


Fig. 3.3 SM-FRET NC titration of –L3L4 TAR DNA with (1mM) EDTA-treated NC ([Mg^{2+}]= 0.2mM in solution).

Binding of NC to TAR hairpins

NC binding properties to nucleic acids has been investigated with various techniques, such as gel-shift assay,(108, 109) fluorescence anisotropy, (109) tryptophan quenching, (110, 111) surface plasmon resonance (SPR) (94) and UV absorbance (112). However, NC binding to TAR DNA and TAR RNA which are the primary reacting nucleic acids involved in the HIV-1 minus-strand transfer step has not been carefully examined. Only few studies on NC binding to TAR RNA have been reported (108, 109) and none on wild type TAR DNA. In previous gel shift studies, TBE (Tris-Borate with 1mM EDTA) running buffer has been routinely used in native gel-shift studies for NC-nucleic acid binding. However, EDTA is known as an effective zinc chelating reagent which can destroy the zinc-finger structure of NC by ejecting the zinc ion out of the protein. In our study, we examine the EDTA effect on binding pattern of NC/TAR in gel

shift assay. TBE buffer and TB (No EDTA) were used as gel running buffer respectively, the results are shown in Fig. 3.4A and 3.4C. In TB buffer, three NC-TAR DNA complex complex bands were observed, whereas in TBE buffer, only one TAR/NC complex band was observed. The results clearly demonstrate that the NC-TAR DNA binding pattern can be dramatically changed due to the presence of EDTA.

EDTA is a strong zinc-binding agent, with a K_d of ~ 10^{-16} M.(113) In addition, EDTA binds zinc much more tightly than other common divalent metals such as Mg^{2+} (K_d ~ 10^{-9} M) and Ca^{2+} (K_d ~ 10^{-11} M),(113) which means that EDTA can selectively and effectively eject zinc out of NC, especially under our experimental conditions where EDTA is large excess compared to NC (<2uM). Furthermore, the releasing rate of zinc ion from NC's zinc fingers has been found to be ~ 3 s^{-1},(114) suggesting that only few minutes exposure to excessive EDTA results in the complete removing of zinc from NC. Since releasing zinc ion from HIV-1 NC induced the unfolding of the protein, (114, 115) running gel-shift assay with TBE buffer (for more than two hours in our experiments) is actually probing the binding of nucleic acids with NC lack of proper folded zinc-finger structure instead of the intact wild-type NC. Our results suggest it is of great importance to keep the integrity of the zinc fingers to its binding functionality. Multiple discrete bands of TAR/NC complex we observed in TB running buffer is distinctly different compared to only one TAR/NC complex band in previous studies using TBE. Considering the fact that zinc-finger is required for NC chaperone activity (Fig. 3.3), the multiple bands we observed might correspond to various secondary structured TAR hairpins associated with NC. The apparent K_d for NC binding to TAR DNA in the absence and in the presence of EDTA are comparable, ~100nM. The results indicate that even in the absence of a folded zinc-finger structure, NC is capable of associating with nucleic acid strongly through electrostatic interaction (2).

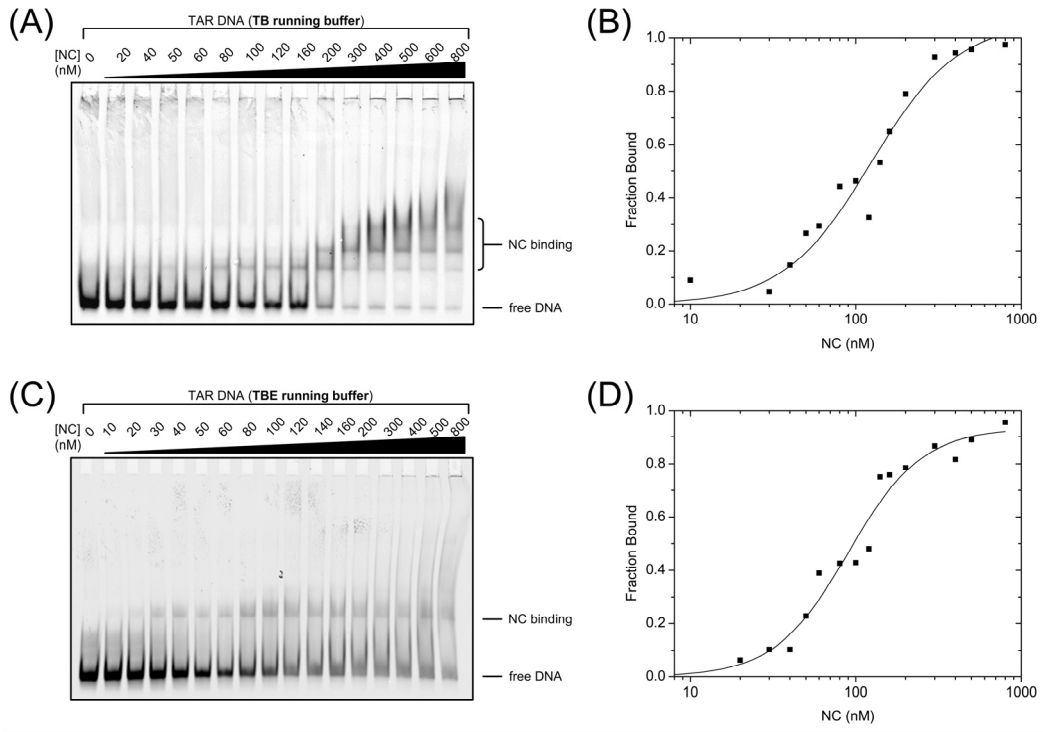


Fig. 3.4. Comparison of gel mobility-shift assay of NC binding to TAR DNA under different gel running buffers. (A) TB (Tris-Borate) and (B) TBE (Tris-Borate-EDTA). 25nM TAR was incubated with NC at various concentrations as indicated for 15 minutes before loading to gel in HEPES buffer (pH 7.3, 40mM NaCl, 0.2mM Mg^{2+}). Native 10% polyacrylamide gel was prepared and run at room temperature. Gel images were taken with the Typhoon gel scanner with excitation at 532nm and emission above 670nm.

NC binding to different secondary structures are also investigated with gel mobility-shift assay. Fig. 3.5 shows the gel-shift data of NC bound to various TAR DNA/RNA mutants, including –L3L4 TAR, –L1L2 TAR, –L1L2L3L4 TAR, cTAR and TAR RNA. –L3L4 TAR, –L1L2 TAR and –L1L2L3L4 TAR (Fig.3.5A-3.5C) all show different binding patterns with TAR DNA (Fig.3.4A), –L3L4 TAR has two binding complexes bands, –L1L2 TAR has only one, and –L1L2L3L4 shows no clear binding complex band. As the number of the internal bulges decreases, the number of the binding binding bands decreases, in other words, more discrete complex bands for NC with those

nucleic acids which have more single-stranded regions. All bulges-deleted mutant of TAR DNA (–L1L2L3L4 TAR) shows no significant binding complex band. Taken together, this data suggest that more discrete binding sites exist for nucleic acids with more NC-meltable regions and provide the evidence that NC favors the binding to single-stranded regions of nucleic acids. This is consistent with the observations from NMR spectroscopy that NC has a strong interaction with loop regions of SL3 and SL2 of HIV-1 genomic RNA Ψ-site.(93, 116) Using other techniques, such as surface plasmon resonance (SPR),(51) fluorescence quenching,(110, 117, 118) nitrocellulose filter RNA binding assay,(119) it also has been shown that NC binds more strongly to single-stranded DNA/RNA than double-stranded regions.

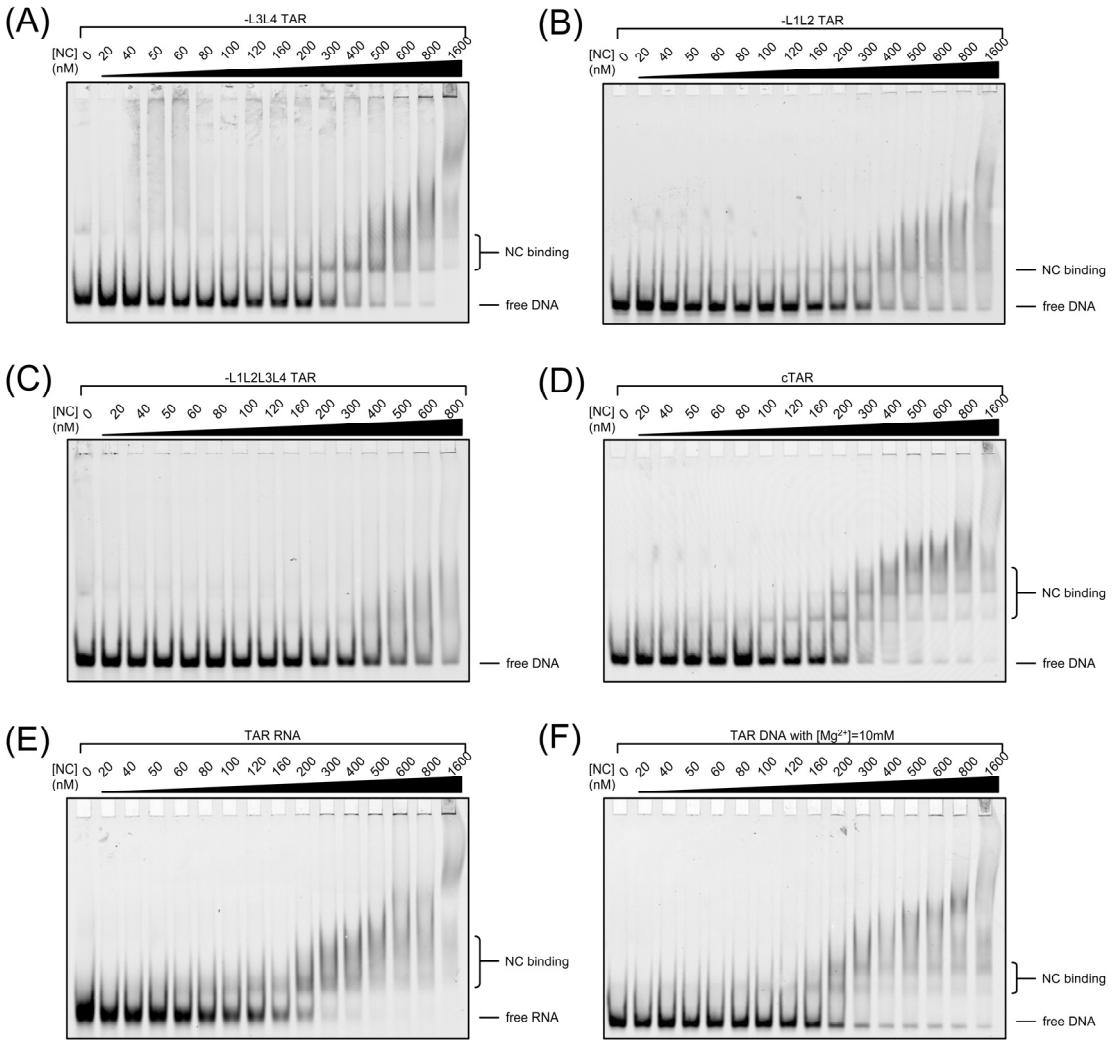


Fig. 3.5 Gel mobility-shift assay of NC binding to various TAR hairpins.

(A) –L3L4 TAR, (B) –L1L2 TAR, (C) –L1L2l3L4 TAR, (D) cTAR, (E) TAR RNA and (F) TAR with 10 mM Mg^{2+} present. 25nM of each nucleic acid was incubated with NC at various concentrations as indicated for 15 mins before loading to gel in HEPES buffer (pH 7.3, 40mM NaCl, 0.2mM Mg^{2+}) (except for (F) with 10mM Mg^{2+}), gels were prepared with 10% polyacrylamide and run at room temperature. The images were taken with Typhoon gel scanner with excitation at 532nm and emission above 670nm.

Table 3.1. Binding parameters of NC binding to various TAR hairpins by gel mobility shift assay.

Nucleic acid	Apparent K_d (nM)	Hill coefficient, n
TAR DNA (in TB)	123 ± 19	1.65± 0.09
TAR DNA (in TBE)	90 ± 8	1.95± 0.05
–L3L4 TAR DNA	135 ± 31	1.56± 0.11
–L1L2 TAR DNA	190 ± 52	1.27± 0.14
–L1L2L3L4 TAR DNA	147 ± 24	1.25± 0.08
TAR DNA ($[Mg^{2+}]$ = 10 mM)	118 ± 23	1.36± 0.09

Despite the difference in secondary structure, TAR hairpins including bulge mutants exhibit apparent NC binding constant, in the range of 100~150nM (Table 3.1). On the other hand, the binding pattern strongly depends on the secondary structures of the TAR DNA hairpins (Fig. 3.5). Clear three bands are observed for wild-type TAR DNA associated with NC. Deletion of the available four bulges results in decreases in the number of the nucleic/NC complex bands. In the absence of EDTA, NC associated with TAR DNA hairpins which possess several single-stranded bulge regions. The strong interaction can hold the complexes in an associated form which migrate down through the gel matrix without effective dissociation. In the presence of EDTA or deletion of all single-stranded bulges (-L1L2L3L4), the bands corresponding to nucleic acid/NC complexes disappear or get expanded. The results indicate that both the zinc-finger and single-stranded regions are required for strong interaction between nucleic acids and NC.

Gel-shift assay for TAR DNA at a high Mg^{2+} concentration (10mM) was also performed in TB gel running buffer (Fig. 3.5F). Mg^{2+} is a competitor of NC in the interaction with nucleic acids. It has been shown that Mg^{2+} has a counter effect of NC by keeping the hairpin structure in a closed form, (12, 13) our NC-concentration dependent SM-FRET measurements also confirmed this (Fig. 3.6). By raising the Mg^{2+}

concentration from 0.2 mM to 10 mM, the equilibrium is driven from partially opened TAR toward closed TAR. Subsequently the gel mobility-shift data (Fig. 3.5F) show that there are less NC-TAR DNA binding complex bands at higher Mg^{2+} concentration, yet similar apparent K_d.

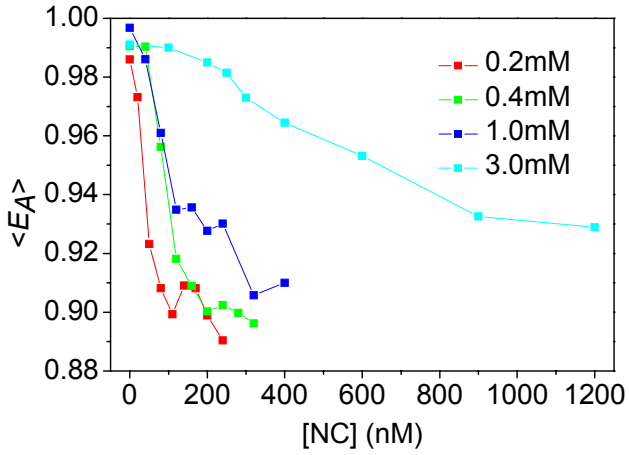


Fig. 3.6 SM-FRET NC titrations of –L3L4 TAR DNA at different Mg^{2+} concentrations (0.2 mM, 0.4 mM, 1.0 mM and 3.0 mM as labeled).

Thermodynamic stability of TAR hairpins secondary structure

Thermodynamic stabilities of various TAR DNA hairpin structures are evaluated using the results from our SM-FRET melting and annealing experiments. This is a qualitative approach to identify the regions with different accessibility by NC or other target oligonucleotides in the presence of NC. The generated color map of the reaction "hot spot" within the TAR hairpins are summarized in Fig. 3.7A.

As shown in Fig. 3.1, a decrease of $<E_A>$ value of ~1 for TAR and –L3L4 to ~0.8 in the presence of NC indicates that their L1L2 bulges are melted by NC. Previous studies have demonstrated that the L1L2 bulges in –L4 and the L1 bulge in –L2L3L4 are also melted in the presence of NC.(12, 13) Therefore, these regions are considered as

"meltale" by NC and coded in red (Fig. 3.7). SM-FRET experiments also show that –L1L2L3L4 remains in a closed conformation in the absence and presence of NC (Fig.3. 1), indicating that all regions in this perfectly base-paired DNA hairpin are not melted. In In addition, previous SM-FRET results on –L4, –L3L4, –L2L3L4 and –L1L2L3L4 have demonstrated the bottom stem region is in closed form, (12, 13) and these regions are coded in blue.

SM-FRET annealing kinetic results also help us identity the reactive regions in the presence of NC. The extent of activity for different regions was characterized by the reaction rate of the annealing of TAR hairpins with various oligonucleotides. For instance, the L1L2 top region accessibility was examined by annealing of TAR DNA to zipper DNA(16). Forward annealing rate of this process is the same as the annealing rate of TAR DNA to cTAR, and similar reaction rate was also observed for –L3L4 annealing with zipper DNA (105). Therefore, the L1L2 region of TAR and –L3L4 is considered as melted and highly active, color-coded in red.

The L3L4 bulge region of TAR was probed by annealing of TAR DNA with a short oligonucleotide which is complementary to L3L4 bulge region of TAR DNA, and the loop DNA oligo (16). The observed reaction rate is ~5 times slower than the annealing of TAR DNA with zipper DNA (16). Assuming the annealing rates of both pathways have the same pre-exponential factor, this suggests that the energy barrier for annealing through loop pathways is ~1 kcal/mol higher than the energy barrier for annealing through zipper pathway. Thus, the L3L4 bulge and loop region in TAR DNA is considered as "softened" and marked in orange. No annealing between –L1L2 and cTAR was observed within our experimental dynamic range, suggesting that the top regions in –L1L2 are not melted by NC (104), therefore coded in blue.

Next, we used mfold to investigate the thermodynamic stability of different regions within TAR hairpins (Fig. 3.7B). Mfold is a well know program for predicting nucleic acid secondary structure (103, 120, 121). For each DNA hairpin, different melted forms were predicted by mfold and their free energies were calculated. To eliminate misfolded states and other sub-optimal states from mfold prediction, constraints have been applied to let the program generate the particular secondary structures we are interested in. For example, TAR DNA, the fully closed form gives the lowest free energy (folding executed in the same salt condition as experiments and at room temperature) of -7.25 kcal/mol; partially open "Y" form (top two bulge regions are melted) has the free energy of -7.21 kcal/mol, which is slightly higher than the closed form; melting form of the bottom loop (L4 and loop region) has -4.94 kcal/mol free energy; melting form of only L3 bulge region open gives -2.06 kcal/mol free energy. In order to investigate the thermodynamic stability of each region, we compared the free energy of different folded secondary structures with the most stable structure (lowest energy form). For wild-type TAR DNA, there is <0.5 kcal/mol energy difference between the "Y" form and the closed form; ~2 kcal/mole energy difference between L4 bulge region open form and closed form; ~ 5 kcal/mol energy difference between L3 bulge region open form and the closed form. A small energy difference indicates this region can be easily open, while a large energy difference means this region needs much more energy to be melted. In other words, the probability of each different folded secondary structure obeys Boltzmann distribution; the folded form with lower energy has a larger population compared to the high energy form. Finally, other TAR DNA mutants were analyzed in the same manner and the whole energetics map was generated.

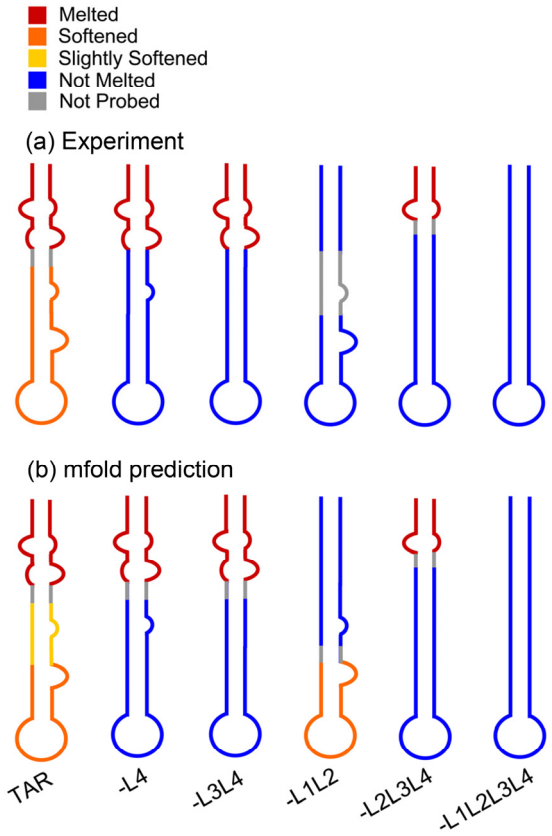


Fig. 3.7 Energetics TAR hairpins in the presence of NC.
(A) Estimated by SM-FRET melting and kinetic annealing experiments. Red: melted by NC and accessible to annealing with zipper DNA oligomer (complementary to L1L2 region of TAR). Orange: accessible to annealing with loop DNA oligomer (complementary to L4 and bottom loop region of TAR) and a short 14mer (complementary to L3L4 region of TAR). Blue: Neither melted nor accessible to annealing with cTAR DNA, zipper DNA/loop DNA. Grey: region not probed. (B) Predicted by mfold. The energy need to melt each region is marked as: Red: <0.5 kcal/mol; Orange: 2 kcal/mol; Yellow: 5 kcal/mol; Blue: > 12 kcal/mol; Grey: not probed.

There is a good agreement between experimental scheme of melted regions and the predicted energy scheme from mfold (Fig. 3.7). This demonstrates that the NC melting ability highly depends on the nucleic acids secondary structure. The presence of multiple internal bulges and loops plays important role in the melting process. The region with lower energy to be opened correlates with the region easily to be melted or accessible for annealing in our experiments. In our thermodynamic stability map, it is

clearly shown that easily open regions are located in the bulge region, loop region or bulges near termini; meanwhile fully base-paired double-stranded region are much more stable and not melted by NC. The presence of these single-stranded regions provides stronger binding sites for NC, at the same time, lowers the energy cost for shifting the conformation to a melting form, which is an important intermediate in annealing reaction of nucleic acid hairpins.

CONCLUSIONS

HIV-1 NC is a multifunctional protein, involving many steps of HIV-1 life cycle (115), especially its chaperone activity is of great importance in nucleic acid secondary structure rearrangement in reverse transcription. Here, we employed SM-FRET approach and gel mobility-shift assay to investigate NC concentration and TAR hairpins secondary structure dependence on NC binding and melting properties. Our SM-FRET results show that the melting of TAR hairpins (and mutants) occurs within a narrow range of NC concentration from 50 nM to 150 nM, with a C_m of ~ 100 nM. Acceleration in the reaction rate of TAR+cTAR annealing falls in the same NC concentration range. Strong correlation between the critical NC concentration for melting and annealing further confirms the reaction mechanism we proposed previously (Chapter 2). Our gel-shift studies on TAR RNA/DNA (and mutants) in the TB buffer first time show the multiple discrete bands for nucleic acids/NC complexes. Although the apparent K_d is similar for all TAR hairpins, secondary structure dependence on gel binding patterns were clearly demonstrated. The thermodynamic stability map of different regions within TAR hairpins was generated based on SM-FRET experimental results and mfold calculation. The easily melted/reactive regions are located at the internal bulges, hairpin loops or bulges near

hairpin termini. This analysis can be potentially used as a qualitative approach for identifying NC melted regions for a wild range of nucleic acid structures.

Chapter 4 A comparative analysis of RNA/protein dynamics for the arginine-rich-binding-motif (ARM) and zinc-finger-binding-motif proteins encoded by HIV-1

INTRODUCTION

The advent of single-molecule fluorescence resonance energy transfer (SM-FRET), with its ability to unravel the complex structural dynamics of biomolecules (4, 5, 11, 122-124), has made it possible to analyze RNA/protein dynamical interactions at the molecular-level with unprecedented specificity, for example obtaining direct information on the secondary structure of key functional, ribonucleoprotein complexes (125-127). Herein we use SM-FRET in vitro to systematically compare and contrast the RNA/protein interactions for two critical motifs for RNA/protein binding in the HIV-1 lifecycle, namely the arginine-rich-binding-motif (ARM) and the zinc-finger-binding-motif. HIV-1 encodes two ARM regulatory proteins, Rev and Tat, and a zinc-finger-binding-motif protein, the multifunctional nucleocapsid (NC) protein. SM-FRET results on the interaction of Tat and NC with a series of key viral RNA sequences give a clearer picture of how these two binding motifs differ with regard to their sequence specificity, impact on RNA secondary structure, and chaperone-like nucleic acid annealing activity.

Tat functions as a highly efficient transcriptional activator of HIV-1 through binding to the transactivating response element (TAR) RNA hairpin (Fig. 4.1) located at the 5' end of the untranslated leader region of the viral mRNA (128-132). Tat is an 86-amino acid protein that contains a cysteine-rich domain, a core region composed of hydrophobic amino acids, a nuclear localization region containing ARM responsible for the RNA-recognition and binding, and a glutamine region. The proposed Tat binding site on TAR RNA, the UCU bulge, is indicated by a red arrow in Fig. 4.1. The 20-amino acid

Tat peptide containing only the core region and the ARM (Fig. 4.1) has been identified to be sufficient for the recognition and specific binding to TAR RNA (133, 134) and is used in the present studies.

In contrast to Tat, NC can bind to various DNA and RNA hairpin structures using its zinc fingers and other structural domains (2). HIV-1 NC is 55 amino acids in length and has two CCHC-type zinc fingers, each of which binds a zinc ion. NC is a multifunctional protein which plays a role in almost every step of the retroviral life cycle, from packaging and assembly to reverse transcription and DNA integration (2). While some NC functions, such as genomic RNA packaging, are believed to involve sequence-specific binding to nucleic acids, NC also displays more general, nonsequence-specific nucleic-acid binding properties (2). In addition to its role as a structural protein that stabilizes the virion, NC also serves as a nucleic acid chaperone that catalyzes the rearrangement of both DNA and RNA into thermodynamically more stable structures and promotes several strand annealing reactions during reverse transcription (135-145), such as the annealing of the structured hairpins, TAR RNA and TAR DNA (Fig. 4.1) to form a fully base-paired duplex in the minus strand transfer step of reverse transcription (25, 46, 135, 137).

The chaperone activity of NC is believed to arise from two main consequences of NC binding to nucleic acids. First, NC lowers the energy barrier for annealing by partially melting the Watson–Crick pairing of the duplex regions due to NC preference for binding to single-stranded bases of both RNA and DNA (22, 145, 146). Second, NC lowers the energy cost of bringing two complementary hairpins together to form encounter complexes by screening the negative charges of the hairpins (27, 147-149).

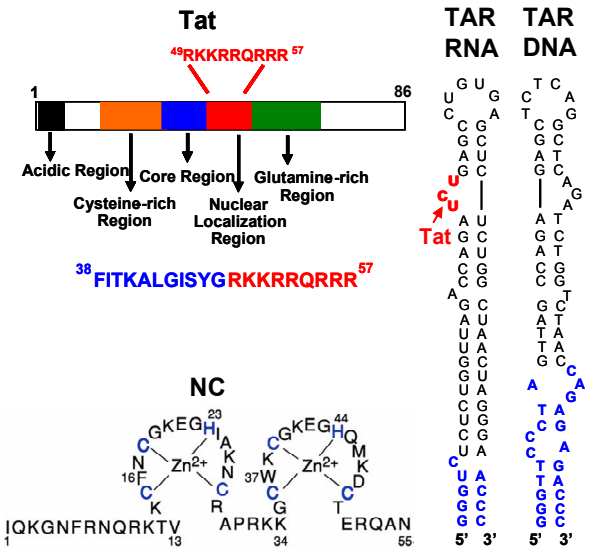


Fig. 4.1 Primary structures of HIV-1 Tat, and nucleocapsid (NC) proteins and secondary structures of HIV-TAR RNA and TAR DNA. NC(1-55) proteins and Tat(38-57) peptide containing the core region and nuclear localization region of the Tat protein were used in the present studies.

We have previously used SM-FRET to identify melted secondary structure intermediates in the NC chaperoned annealing of TAR RNA and TAR DNA (136, 138-141). The blue regions of the secondary structures in Fig. 4.1 for TAR RNA and TAR DNA have been shown to be fully melted by NC at concentrations > 300 nM. In this chapter, we see how these TAR DNA/TAR RNA annealing reactions are inhibited by sequence specific binding of Tat. The inhibition experiments are analogous to previous experiments on how the small nucleic acid-binding molecule argininamide inhibits annealing of TAR DNA and complementary cTAR DNA due to a decrease in hairpin flexibility associated with argininamide binding (150). In this study, direct competitive binding and annealing SM-FRET experiments with pairs of these proteins, Tat/NC and various RNA sequences are used to study how viral proteins of the ARM and zinc-finger motif type interact at the secondary structure level by binding simultaneously to the same RNA sequence. In principle, the interactions of the ARM and zinc-finger proteins with

RNA substrates and the resulting chaperone activities may occur in vivo, when both types of proteins are present in the same vicinity or even in the same ribonucleoprotein complex of a HIV-1 infected cell (151, 152).

MATERIALS AND METHODS

Sample preparation

RNA and DNA oligonucleotides containing appropriate dye-labeling and biotin-functionalization were purchased from Trilink BioTechnologies (San Diego, CA) and were purified by the supplier using RNase-free HPLC. All the molecular constructs used in the present studies are listed in Tables 4.1. The oligonucleotides were labeled with either Cy3 (donor dye) or Cy5 (acceptor dye). Biotin functionalization was added to the nucleotides for the immobilization of the molecules on coverslips through biotin-streptavidin interactions. To prevent the undesirable G residue quenching effects, a UUU or TTT overhang was added to the RNA or DNA sequences. The HIV-1 NC protein for these experiments was prepared by solid-phase synthesis as described previously (136, 138). The Tat(38-57) peptide was synthesized using Rink amide resin on a Applied Biosystems Model 433A peptide synthesizer and standard Fmoc chemistry. The peptide was cleaved from the resin and purified by HPLC on a C4 reverse-phase column using an acetonitrile gradient in 0.1% trifluoroacetic acid (TFA).

Gel mobility shift assay

The experimental details of the gel-shift assays are summarized in Table 3. Typically, 25 nM Cy3-labeled or Cy5-labeled oligonucleotides were titrated with proteins, and free nucleic acids and nucleic acid-protein complexes were resolved on polyacrylamide gels and imaged using a Typhoon Molecular Imager (Amersham

Biosciences, Piscataway, NJ) under Cy3-or Cy5-flourescence detection mode. The relative integrated intensities of each band on the gels were quantified by QuantityOne software (Bio-Rad Laboratory, Hercules, CA). The binding curves were constructed by plotting the fraction of RNA/DNA bound with proteins (θ) (total bound RNA/DNA in all bands divided by the sum of bound and unbound RNA/DNA) as a function of protein concentration. The apparent dissociation constants (K_D) for protein binding to RNAs can be obtained by fitting the data points with the following equation:

$$K_D{}^n = \frac{(1-\theta)(P_0 - nR_0\theta)^n}{\theta}. \qquad (1)$$

R_0 is the total concentration of RNA added in each lane, which is 25 nM. P_0 is the total concentration of proteins added in each lane. n is a coefficient related to the protein binding stoichiometry. K_D and n values for each protein binding case were obtained by curve fitting to achieve the minimized difference between the experimentally measured θ and calculated θ.

Flow system for oligonucleotide annealing reactions

The annealing reactions were carried out in a home-built flow cell (136, 140). Typically, the Cy5-labeled oligonucleotides, buffer solutions, and protein solutions were selectively flowed into the flow cell to react with the Cy3-labeled oligonucleotides immobilized on the coverslip surface. The commercial coverslips (Fisher Scientific) were cleaned with piranha (sulfuric acid : hydrogen peroxide, 7:3) and then treated with Vectabond/acetone 1% w/v solutions (Vector Laboratories, Burlingame, CA) for 5 minutes. Each coverslip was subsequently PEGylated and biotinylated, after which a reaction chamber with inlet and outlet ports (Nanoport, Upchurch Scientific, Oak Harbor, WA) was assembled. The details of the chamber assembly process have been described in earlier work (136). Each chamber was treated with streptavidin (Molecular Probes,

Eugene, OR; 0.2 mg/mL in 25 mM HEPES buffer), followed by the immobilization of the biotin-functionalized oligonucleotides. The biotin-functionalized oligonucleotides were renatured by incubating 500 nM oligonucleotides in HEPES buffer (25 mM HEPES, pH 7.3, 40 mM NaCl) for 3 minutes at 80°C, 5 minutes at 60°C, and 10 minutes at 0°C. Then the oligonucleotide solutions were further diluted to a final concentration of of 500 pM in HEPES buffer containing 10 mM MgCl2. The immobilization of the oligonucleotides on the coverslips was accomplished by incubating renatured oligonucleotides in the chamber over the streptavidin-functionalized coverslip surface several times. All the annealing reactions were carried out at room temperature in protein-binding buffer in an oxygen scavenger system (153) containing β-D(+)glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL and catalase 0.02 mg/mL (Roche Applied Science, Hague Road, IN). 1% v/v 2-mercaptoethanol was added in the mixtures to eliminate the blinking of molecules and generate longer-lasting single-molecule fluorescence signals (35). The protein-binding buffers were composed of 25 mM HEPES, pH 7.3, 40 mM NaCl, and 0.2 mM $MgCl_2$ (for TAR-related annealing experiments).

Data collection and analysis

A home-built sample scanning confocal optical/data collection system based on a a Zeiss inverted microscope (140) was used in these SM-FRET experiments. The sample flow-cell was scanned by a Queensgate X,Y scanning stage (NPW-XY-100A, Queensgate, Torquay, U.K.). A high numerical aperture, oil immersion microscope objective (Zeiss Fluar, 100×, NA 1.3) was used for excitation and signal collection. The donor and acceptor fluorescence were separated by a dichroic beam splitter (Chroma 630 DCXR, Chroma Tech., VT) into two beams, and each was detected by an avalanche

photodiode (APD) (Perkin-Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, QC, Canada). SM-FRET data were collected synchronously through separate detection channels for Cy3 and Cy5 fluorescence intensities.

The annealing reactions were initiated by flowing a freshly prepared protein solution plus Cy5-labeled oligonucleotides into the flowing sample cell. The SM-FRET images were acquired at several times, t, during the course of annealing reactions and in-house software was used to find molecules, calibrate stage-drift and crosstalk, subtract image background, and calculate SM-FRET efficiency for each molecule. The corrected donor and acceptor intensities, $I_D(t)$ and $I_A(t)$, respectively, were used to calculate the time trajectory of the apparent FRET efficiency, EA(t), according to:

$$E_A(t) = \frac{I_A(t)}{I_D(t) + I_A(t)} . \qquad (2)$$

$E_A(t)$ is related to the actual FRET efficiency, $E_{FRET}(t)$, by the inclusion of the dye quantum efficiencies, Φ_i, and detector quantum efficiencies, η_i, according to:

$$E_{FRET}(t) = \frac{I_A(t)}{I_A(t) + I_D(t)\frac{\phi_A \eta_A}{\phi_D \eta_D}} . \qquad (3)$$

In the case of the current experimental setup, it was determined that $E_A(t) \approx E_{FRET}(t)$. The collected donor and acceptor signals were corrected for background emission/noise and donor/acceptor crosstalk due to signal leakage as previously described (136, 139).

RESULTS AND DISCUSSION

This section is focused on a comparison of how NC and Tat separately and in combination interact with the TAR oligomers that are summarized in Table 4.1. We are

especially concerned with how the presence of Tat inhibits the NC-chaperoned annealing reaction between TAR DNA and complementary TAR RNA or alternatively annealing between TAR DNA and complementary cTAR DNA (136, 138-141), which is portrayed in the bottom of Table 4.1. The various oligomers have been appropriately labeled with Cy3 (donor dye), Cy5 (acceptor dye) and biotin for immobilization of the oligomers on the coverslip. The experiments include gel mobility shift assays in order to establish specific and non-specific binding affinities for these nucleic acid binding proteins and set of oligomers, see Fig. 4.2 for typical examples. This is followed by SM-FRET experiments on the rates and yields of the NC-chaperoned annealing reactions in the presence and absence of Tat.

Oligonucleotides	Primary sequences
TAR DNA	5'- Cy3-TGGGTTCCCTAGTTAGCCAGAGA GCTCT(biotin)CAGGCAGATCTGGTCTAA CCAGAGAGACCCTTT -3'
TAR RNA	5'- GGGUCUCUCUGGUUAGACCAGAUCU GAGCCUGAGAGCUCUCUGGCUAACUAG GGAACCCUUUU-Cy5 -3'
cTAR DNA	5'-TTTTGGGTCTCTCTGGTTAGACCAG ATCTGAGCCTGAGAGCTCTCTGGCTAA CTAGGGAACCCT-Cy5 -3'

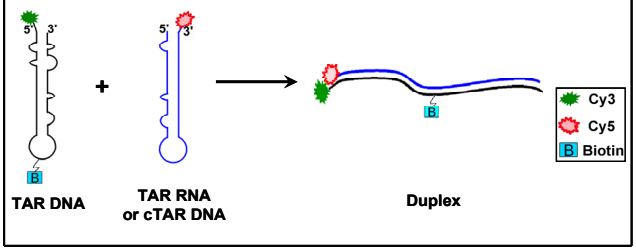


Table 4.1. Primary sequences of TAR-related oligonucleotides and schematics illustration of the annealing reactions.

As shown in Table 4.2 and Fig. 4.2A, Tat binds strongly to its target sequence on TAR RNA with an apparent dissociation constant (K_D) of ~ 15 nM but not to TAR DNA or the other non-TAR sequences, consistent with ARM binding motif assignment for this

peptide (see Introduction). It is apparent from Figure 2A that two copies of Tat can bind on one TAR RNA hairpin at higher protein concentration. In contrast to Tat, NC binds strongly to both TAR DNA and TAR RNA with K_D of < 100 nM (Fig. 4. 2B, Fig. 4.2C, and Table 4.2). The "smearing" of the bands for NC binding is consistent with previous reports that multiple copies of NC bind with a "footprint" of one NC per ~8 bases pairs without strong sequence specificity to hairpin and fully duplexed DNA and RNA (147). It is important to note that while NC is known to melt structured DNA and RNA hairpins due to its binding preference for single-stranded nucleic acids, melting is not required for strong binding. Consistent with these trends we observe strong binding (K_D < 200 nm) to fully duplexed DNA (TAR-cTAR DNA duplex, in HEPES buffer) which should not be melted by NC. We also observed strong binding of NC to RNA with K_D ~ 100 nM (see Fig. 3.4 in Chapter 3) in the presence of ethylenediaminetetraacetic acid (EDTA), a chelating ligand that strongly interacts with zinc ions and therefore can be used to effectively extract the zinc ions out of the zinc fingers of NC (154). In the absence of zinc, the zinc fingers can no longer maintain their properly folded structures that are crucial to the local melting of the duplex regions of the hairpin. These results are consistent with a non-sequence specific, electrostatic binding of NC to DNAs and RNAs, being an important if not dominant binding mode under these conditions. This is consistent with the conclusion that NC can fold into different conformations to accommodate productive electrostatic interactions with the RNA substrate even in the absence of Zn^{2+} binding (154, 155).

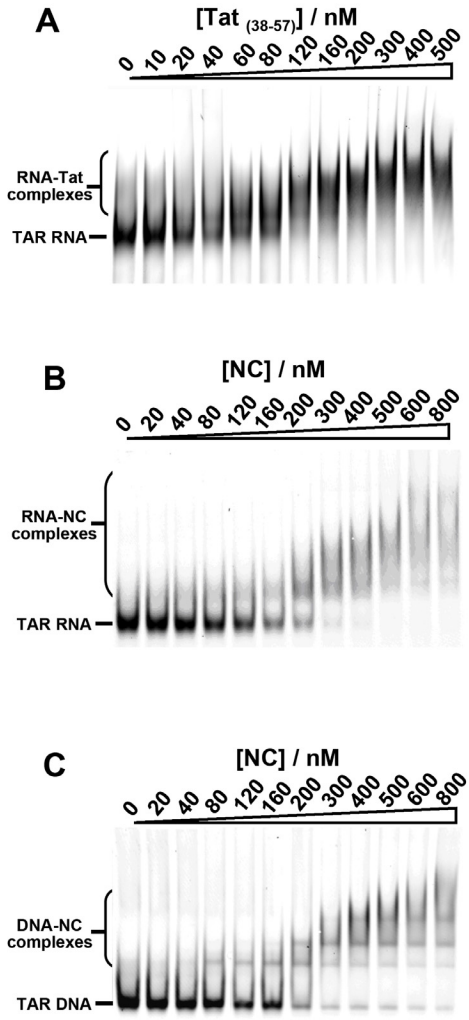


Fig. 4.2 Electrophoresis gel-shift assay of (A) Tat binding to TAR RNA, (B) NC binding to TAR RNA, and (C) NC binding to TAR DNA.

It is interesting to examine whether the observed Tat and NC binding affinity pattern is reflected in how these proteins affect TAR annealing reactions. Typical SM-FRET annealing data for the NC induced annealing reaction of immobilized TAR DNA hairpins with 5 nM TAR RNA and 800 nm NC in a flowing buffered solution are shown in Fig. 4.3. Each single molecule FRET trajectory $E_{FRET}(t)$ in the upper panel corresponds to a single immobilized oligomer for which the FRET value EFRET

switches from zero to unity when the annealing reaction occurs. The lower panel in Fig. 4.3 is the fraction of annealed TAR DNA hairpins as a function of reaction time. We have extensively described how data of this type can be analyzed to obtain accurate bimolecular annealing rate constants corresponding to the reaction in the bottom panel in Table 4.1 (140, 141).

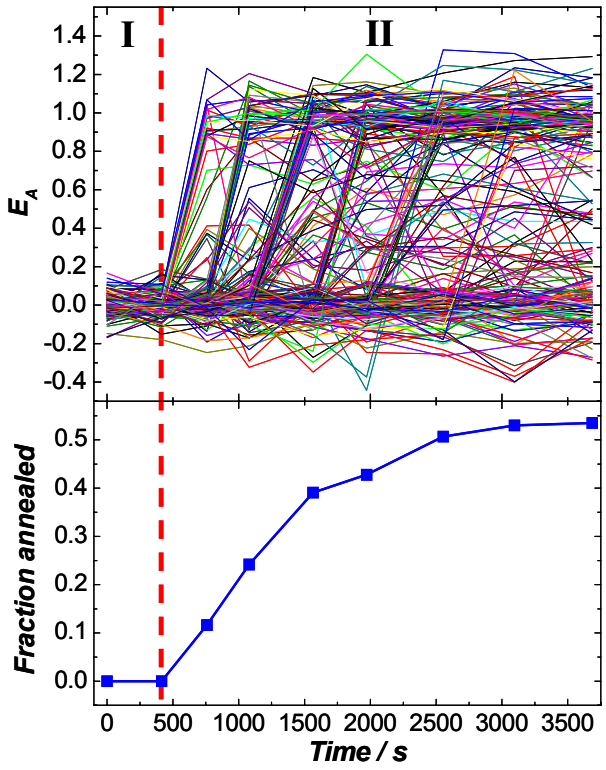


Fig. 4.3. SM-FRET trajectories of 215 molecules found in a 30 μm × 30 μm region (upper panel) and the fraction of annealed TAR DNA hairpins as a function of reaction time (lower panel) during the annealing of Cy5-labeled TAR RNA to immobilized Cy3-labeled TAR DNA. During time epoch I, 5 nM TAR RNA was flowed into the reaction chamber. During time epoch II, 5 nm TAR RNA and 800 nm NC were co-flowed into the reaction chamber.

The FRET histograms in Fig. 4.4, which were recorded at long times (e.g., 1 h) after the reagents were mixed for an ensemble of hairpins, give a clear picture of the

impact of Tat and NC on the annealing reaction. Under these conditions equilibrium mixture strongly favors the fully annealed duplex. In the absence of either NC or Tat the annealing is too slow to observe on this time scale and only the EFRET(t) ~ 0 peak in the histogram is present (Fig. 4.4A). With the 800 nM NC added to the TAR RNA/buffer solution annealing becomes rapid enough to observe a significant EFRET ~ 1 peak in this time window (Fig. 4.4B). The lack of 100% annealing is due primarily to unreactive, misfolded or mislabeled TAR RNA hairpins (140). In contrast to the chaperone-like activity of NC, Tat (Fig. 4.4C) shows no evidence of catalyzing the annealing reaction. Furthermore, Tat is observed to strongly inhibit the chaperone-like activity of NC as shown in Figure 4D. As previously shown, the corresponding DNA only annealing reaction (i.e. TAR DNA + cTAR DNA) is also extremely slow without added NC, but becomes strongly catalyzed with NC present (Fig. 4.5A and 5B) leading to nearly 100% annealing yield within 1 h. For the TAR DNA + cTAR DNA case, Tat alone is unable to to catalyze the annealing (Fig. 4.5C) and adding Tat to the NC reaction has no measurable inhibitory effect (Fig. 4.5D). This strongly suggests that Tat specifically binds to the TAR RNA hairpin, and only through this specific binding is the inhibition of NC-chaperoned annealing accomplished.

The TAR DNA + TAR RNA annealing results that are summarized in the top row of Table 4.3 can be rationalized by considering the previously proposed mechanism for NC chaperone activity (136). It has been hypothesized that NC partially melts both reactant hairpins allowing for rapid nucleation of annealing initiated by base-pairing of short single stranded regions. The melting process has been directly observed by investigating the SM-FRET of dual labeled TAR hairpins with NC present (136, 139). NC induced partial melting has furthermore been assigned to selective stabilization of melted secondary structures of TAR DNA and TAR RNA by NC binding, resulting from

NC's binding preference for single stranded regions. It has been proposed that the nucleation of annealing occurs right at the locally melted regions in an encounter complex that is comprised of partially melted TAR DNA and TAR RNA or cTAR DNA hairpins associated with multiple copies of NC proteins (138, 141). The inhibitory mechanism of Tat on TAR DNA and TAR RNA annealing may simply be due to Tat's ability to block the binding of NC to TAR RNA, thus suppressing NC induced melting of TAR RNA.

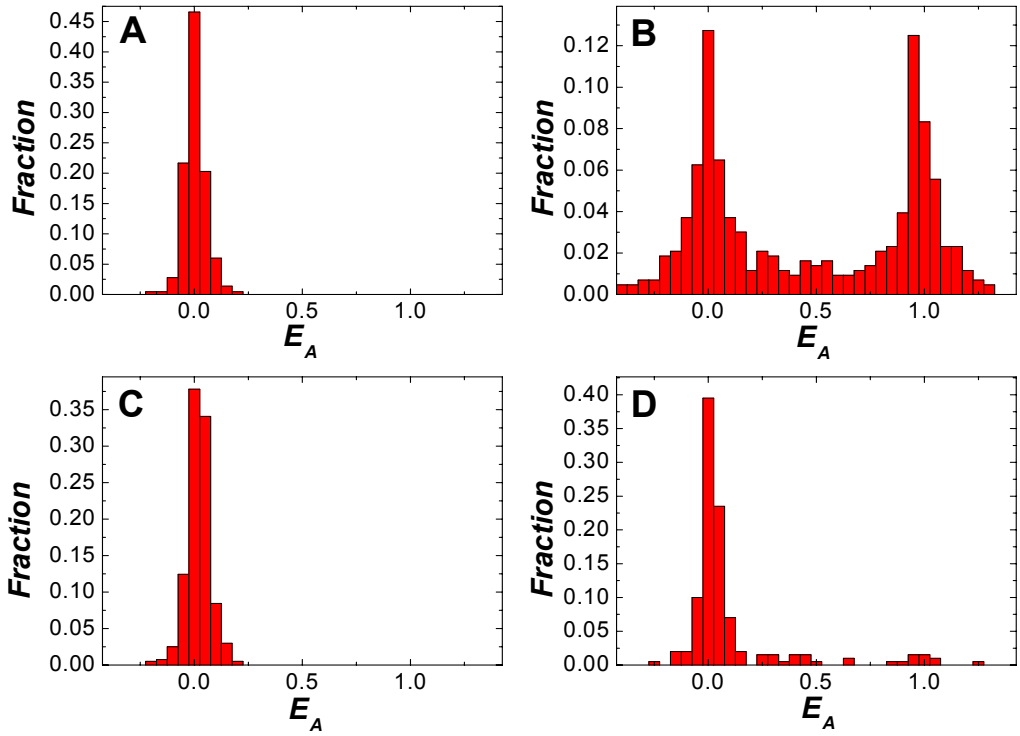


Fig. 4.4 FRET histograms obtained from SM-FRET measurements after flowing in 5 nM Cy5-labeled TAR RNA over immobilized Cy3-labeled TAR DNA for 1 h in the presence of (A) no proteins, (B) 800 nM NC, (C) 500 nM Tat(38-57), and (D) 500 nM Tat(38-57) and 800 nM NC.

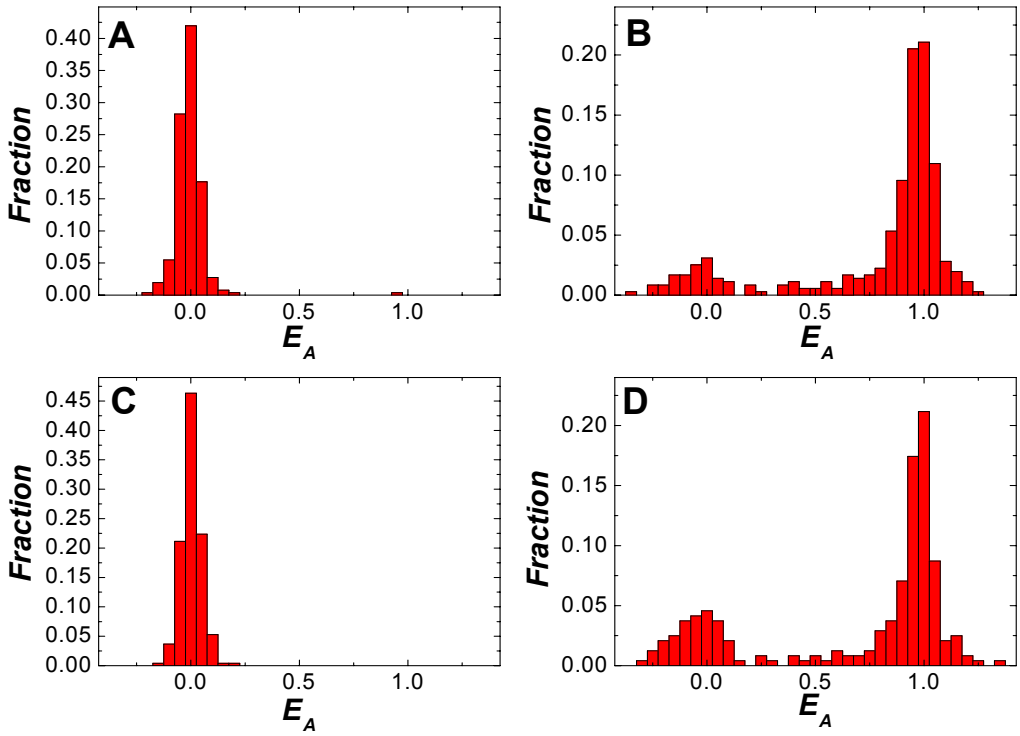


Fig. 4.5 FRET histograms obtained from SM-FRET measurements after flowing in 5 nM Cy5-labeled cTAR DNA over immobilized Cy3-labeled TAR DNA for 45 min in the presence of (A) no proteins, (B) 800 nM NC, (C) 500 nM Tat(38-57), and (D) 500 nM Tat(38-57) and 800 nM NC.

The hypothesis that Tat binding to TAR RNA effectively suppresses partial melting is consistent with the evidence that Tat binds to a secondary structure with the base pairs intact that are adjacent to the red bases in Fig. 4.1 for TAR RNA. Tat's preference for non-melted secondary structures of TAR RNA is also probably a factor in its lack of chaperone activity for TAR RNA+ TAR DNA annealing. Since Tat only binds strongly to TAR RNA, not cTAR DNA (Table 4.2), it is not surprising that Tat has no inhibitory effect on TAR DNA + cTAR DNA annealing, as shown in Fig.4.5D.

Table 4.2. Apparent disorientation constants (K_D) determined by gel-shift assays.

Protein or peptide / peptide / Oligonucleotides	Tat[a]	NC[b]
TAR RNA	15 nM	50 nM
TAR DNA	-----	55 nM
cTAR DNA	-----	120 nM

----- indicates that no obvious protein binding observed in the protein concentration ranging from 0 to 800 nM in these gel-shift assays.

[a] The gel-shift assays for Tat-binding was performed by incubating oligonucleotides with Tat(38-57) peptide at room temperature in 20 μl of binding mixtures containing 20mM Tris base (pH7.5), 100mM NaCl, 10mM DTT, 0.1% Triton-100, 0.2mM Mg^{2+}. The gel-running buffer was 0.5X TBE.

[b] The gel-shift assays for NC-binding was performed by incubating oligonucleotides with NC at room temperature in 20 μl of binding mixtures containing 25 mM HEPES (pH = 7.3), 40 mM NaCl, 0.2 mM $MgCl_2$, and 10 % glycerol. The gel-running buffer was 0.5X TB (Tris-Borate).

Table 4.3. Chaperone and inhibition assays for HIV-1 RNA binding proteins or peptides

Annealing reaction	NC chaperone activity	Inhibiter of NC induced annealing	These proteins have no chaperone activity	Specific binding
TAR DNA + TAR RNA	Yes	Tat	Tat	Tat binding to TAR RNA
TAR DNA + cTAR DNA	Yes	None	Tat	None

CONCLUSION

We have presented a comparative study on how the binding of two families of HIV-1 viral proteins to viral RNA hairpins locally changes the secondary structures of the RNAs using a SM-FRET approach. ARM proteins (Tat) and a zinc finger protein

(NC) use different RNA binding strategies to recognize and interact with RNA, giving rise to very different changes in the RNA's secondary structure upon protein binding. Our SM-FRET results clearly reveal that NC locally melts TAR RNA hairpin, while Tat strengthens the hairpin structures through specific binding interactions. Competition experiments show that the presence of Tat can effectively inhibit the NC binding-induced local melting of TAR RNA hairpin. This data suggests that in HIV-1, the arginine-rich motifs of Tat help stabilize the RNA hairpin structures, which likely inhibits the local melting of the hairpins induced by NC.

Previous studies have suggested that HIV-1 Tat can promote nucleic acid restructuring reactions, including tRNA primer annealing onto the primer binding site (156), as well as complementary TAR DNA hairpin annealing (157). In contrast to these previous studies, we see no evidence of nucleic acid chaperone activity by Tat38-57 using the RNA/DNA oligonucleotides investigated here. It is known that tRNA primer annealing does not require the zinc-finger structures of NC and can be facilitated by polyLysine (158). Thus, it is not too surprising that Tat, which contains a basic ARM, can also facilitate tRNA primer annealing (156). It is also known that NC is a general nucleic acid chaperone, which can catalyze many nucleic acid rearrangements throughout the HIV lifecycle, whereas Tat functions via specific nucleic acid binding interactions. Interestingly, NC has also been shown to migrate from the cytoplasm to the nucleus, where it has been proposed to play a role in inducing HIV-1 early mRNA expression prior to high level expression induced by Tat (151, 152). The competition binding data presented here suggest that the specific nucleic acid binding interactions of Tat can effectively compete with NC and thus, can regulate critical events such as viral RNA transcription and mRNA nuclear export during the HIV lifecycle despite the presence of a robust nucleic acid binding/chaperone protein such as NC.

Chapter 5: Conformational change of single-stranded DNA by HIV-1 nucleocapsid protein

INTRODUCTION

HIV-1 nucleocapsid protein (NC) is a basic viral protein containing 55 amino acids and two hightly-conserved CCHC-type zinc fingers (2). NC is a multifunctional protein which plays important role in almost every step of the HIV-1 life cycle, from reverse transcription and DNA integration to viral particle packaging and assembly (2). For instance, NC is known as a nucleic acid chaperon, unwinding a cellular tRNA and catalyzing the hybridization of the tRNA to the primer binding site on viral genome (2). After the initiation of the reverse transcription, NC destabilizes the folded transactivation response element (TAR) and facilitates the stand transfer process (136, 138-141). In the mature HIV-1 virion, NC is complexed with genomic RNA and is responsible for stabilization of the two genomic RNA molecules (3). Through all these steps, NC plays its function mainly through binding and interaction with single-stranded or single-stranded regions of nucleic acids. While some of NC functions, such as genomic RNA packaging, are believed to involve sequence-specific recognition and binding to nucleic acids, NC also plays more general, non-sequence-specific nucleic acids binding properties (2).

Many attempts have been put to investigate the structures and binding properties of NC interacting with single-stranded RNA and DNA. Experiments with NC and short oligonucleotides demonstrated that around five bases are required for stable NC binding (51, 159, 160). More results have shown that NC binds to the alternating (TG)

sequence better than others (51, 94). Other studies involving the sequence from HIV-1 genome, which is stem-loop structured has also confirmed the preference of NC for GXG-containing single-stranded sequences (161, 162). NMR structures on several NC/nucleic acids complexes clearly reveal that G residue plays important in NC binding, through the interaction with Trp37 residue of the second zinc-finger of NC (93, 162, 163). Moreover, it has been shown that more than one NC can interact with one (TG)4 molecule (94). However, there is limited data on the conformation of these single-stranded nucleic acid chains associated with NC. Previous data on the structures of the NC/nucleic acid complexes involves relatively short oligonucleotides (51, 94, 162, 164), where only one NC molecule can bind for most of the cases. Therefore, it is important to explore NC's effect on relatively longer single-stranded nucleic acids chain. In principle, the interactions of NC with single-stranded DNA/RNA (~20 nucleotides) may occur in vivo, considering the complex folded genomic RNA and possibly various ssDNA/RNA (chains and regions) which are produced along the HIV-1 replication cycle (2, 165) .

Here, we employ a single molecule fluorescence resonance energy transfer (SM-(SM-FRET) approach to investigate NC-induce conformational change of single-stranded DNA (ssDNA) as well as the NC/nucleic acid complex conformational dynamics at the millisecond to minute time scale. SM-FRET is uniquely capable of unraveling the complex structural dynamics of biomolecules (4, 5, 11, 122-124). To ensure stable NC binding to the ssDNA, we used repeated (TG) sequence with different lengths. Additionally, we compare the binding properties of NC on repeated (TG) sequence to the poly T sequence with the same length to investigate NC's binding preference for specific nucleic acid sequence. Single-stranded DNA chain is immobilized on surface through hybridization with an anchor DNA. Cy3 and Cy5 dye pairs are employed here to report the separation between two ends of the ssDNA chain (Fig. 5.1). We examine the NC

binding and induced conformational change of ssDNA at various protein concentrations, ionic condition, in the absence/presence of competitive short oligonucleotides in solution. These multi-component binding and dissociation experiments of NC/ssDNA interaction allow up to directly probe the stability and dynamics of these ribonucleoprotein complexes.

MATERIALS AND METHODS

Oligodeoxynucleotides

All the DNA oligonucleotides containing appropriate dye-labeling and biotin-functionalization were purchased from Integrated DNA Technologies (Coralville, IA) and were purified by the supplier using RNase-free HPLC. All the sequences of DNA oligonucleotides used in the present studies are listed in Table 5.1. The short anchor oligonucleotide were labeled with Cy3 (donor dye) at its 3' terminus and longer oligos containing single-stranded region of interest were labeled with Cy5 (acceptor dye) at their 5' terminus. Biotin functionalization was added to the anchor oligonucleotides for the immobilization of the molecules on coverslips through biotin-streptavidin interactions.

Sample preparation

Both of the anchor oligo and Cy5-labeled long oligo were renatured respectively by incubating 1uM DNA in HEPES buffer (25 mM HEPES, pH 7.3, 40 mM NaCl) for 2.5 minutes at 80°C, 2.5 minutes at 60°C, and 5 minutes at 0°C. Then two oligo solutions were mixed at the ratio of 1:3 (Cy3-anchor: Cy5-oligo) with 40mM MgCl2 and incubated for 30 min at 37°C for annealing. Next the hybridized ssDNA solution was further diluted to a final concentration of 1 nM in HEPES buffer containing 10 mM $MgCl_2$. The immobilization of ssDNA construct on the coverslips was accomplished by incubating

oligonucleotides in the chamber over the streptavidin-functionalized coverslip surface several times. HIV-1 NC protein for these experiments was prepared by solid-phase synthesis as described previously.(136, 138) Poly-L-Lysine (MW>15,000) was purchased from Sigma (Sigma-Aldrich, St. Louis, MO).

Flow system for controlling buffer solution

During the experiment, various buffer solutions were carried out in a home-built flow cell (136, 140). Various buffer solution containing NC protein, high concentration $MgCl_2$, and non-complementary short DNA oligonucleotides were delivered by separated syringe pumps. All the reactions were carried out at room temperature in buffer containing an oxygen scavenger system(153) which consists β-D(+)glucose 3% w/v (Sigma-Aldrich, St. Louis, MO), glucose oxidase 0.1 mg/mL and catalase 0.02 mg/mL (Roche Applied Science, Hague Road, IN). All buffer contained 25 mM HEPES (pH = 7.3), 40 mM NaCl, 1uM Zn acetate and 0.2mM $MgCl_2$.

Data collection and analysis

A home-built sample scanning confocal optical/data collection system based on a Zeiss inverted microscope (140) was used in these SM-FRET experiments. The sample flow-cell was scanned by a Queensgate X,Y scanning stage (NPW-XY-100A, Queensgate, Torquay, U.K.). A high numerical aperture, oil immersion microscope objective (Zeiss Fluar, 100×, NA 1.3) was used for excitation (514nm) and signal collection. The donor and acceptor fluorescence were separated by a dichroic beam splitter (Chroma 630 DCXR, Chroma Tech., VT) into two beams, and each was detected by an avalanche photodiode (APD) (Perkin-Elmer Optoelectronics SPCM-AQR-15, Vaudreuil, QC, Canada). SM-FRET data were collected as described previously (chapter 2).

RESULTS AND DISCUSSION

Fig. 5.1A shows a schematic of the DNA construct used in our experiments. Each DNA contains an 18 basepair duplex DNA possessing a 3' single-stranded region of varying sequence, d(T)n and d(TG)n. The donor fluorophore, Cy3 was covalently attached at the 5' end of the anchor strand. The acceptor fluorophore, Cy5 was covalently attached at the 3' end of the single-stranded tail. The whole DNA molecule including the fluorophores is denoted as (TG)n and (T)n in the content. This construct allow us to probe the end-to-end distance of the single-stranded tail, while maintaining the similar local environment of the fluorophores for all DNA oligomers.

TABLE 5.1 sequences of oligonucleotides employed in the experiment

Name	Sequence
Anchor oligo	5'-Cy3-TTCCTCGCTGCCGTCGCCT-3'-Biotin
$(TG)_{10}$	5'AGGCGACGGCAGCGAGGAATGTGTGTGTGTGTGTGTGTGTT TT-Cy5-3'
T_{20}	5'AGGCGACGGCAGCGAGGAATTTTTTTTTTTTTTTTTTTTT -Cy5-3'
$(TG)_{20}$	5'AGGCGACGGCAGCGAGGAATGTGTGTGTGTGTGTGTGTG TGTGTGTGTGTGTGTGTGTTTT-Cy5-3'

As illustrated in Fig. 5.1A, a biotinylated dsDNA with a 3' ssDNA tail is immobilized on surface. The donor and acceptor attached report the end-to-end distance through FRET. The ssDNA itself is flexible (persistence length 1.5~3 nm), and its conformational fluctuation is averaged out on the much faster time scale than our measurement resolution. For the $d(TG)_{10}$, two dyes are in relatively far away position due to the repulsion of negatively charged phosphate backbone, giving relatively medium FRET efficiency ~0.50±0.08 (Fig. 1B white histogram) in our standard buffer condition.

Addition of 400 nM NC results in a higher FRET efficiency ~ 0.72±0.09 (Fig.5.1B red histogram). Due to the binding of NC, a basic protein in our solution condition which neutralizes some of the negative charge of DNA backbone leads to conformational change of ssDNA.

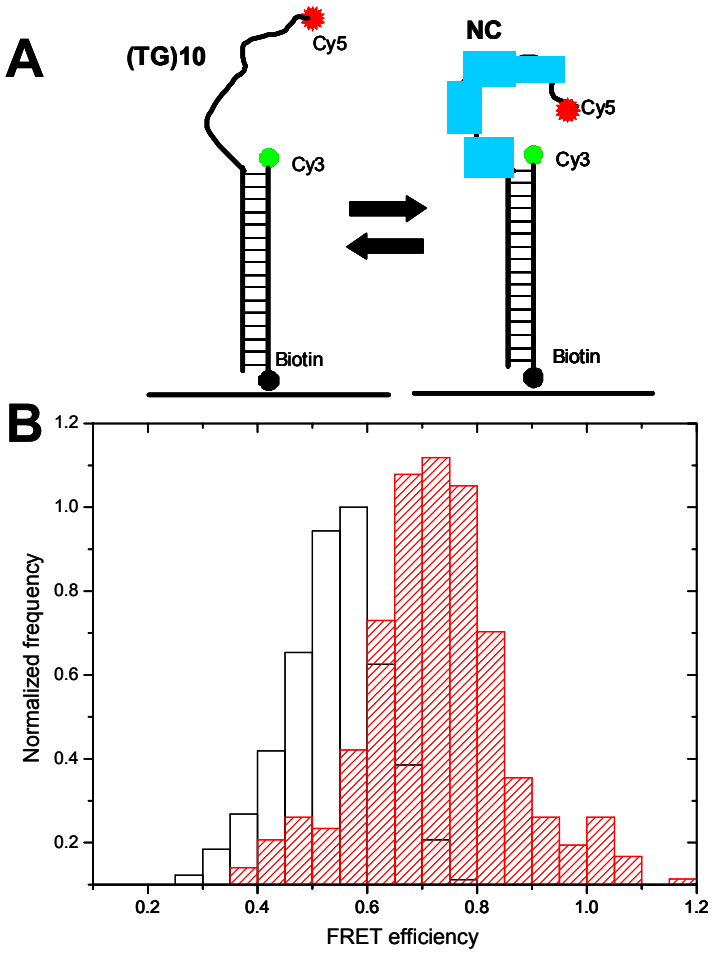


Fig.5.1 NC binding on short ssDNA $(TG)_{10}$
(A) A double stranded DNA with a 3'-$(TG)_{10}$ single stranded DNA tail is immobilized on PEG surface. FRET between donor (Cy3) and acceptor (Cy5) labels reports the changes in the averaged end to end distance of the ssDNA.
(B) FRET histogram of ssDNA $(TG)_{10}$ in buffer only (white) and with 400nM NC (red). Low FRET (Ea~0.55) is from ssDNA only and high FRET (Ea~0.72) is with NC.

Previous study using surface Plasmon resonance (SPR), tryptophan fluorescence quenching (TFQ) and fluorescence anisotropy (FA) shows that the K_d for binding a single d(TG)4 molecule to NC is only 5nM in 150mM NaCl (94). NC has been shown to have a high affinity to the repeating sequence d(TG)n . Here we also investigate the NC concentration effect on $(TG)_{10}$ conformational change. We used extremely low concentration of NC 1nM in the experiment, and we observed a small fraction of molecule exhibiting FRET efficiency increase (Fig. 5.2). Each single-molecule data was collected and ~300 molecules were used to build FRET ensemble histogram. Two clear separated peaks were resolved with FERT peak at ~0.5 and ~0.75 respectively. The lower FRET shows the similar value as what we observed under buffer only condition. While the higher FRET peak shows similar value as the condition in the presence of 400 nM of NC. The higher FRET peak represents those DNA molecules having sufficient amount of NC which fold DNA into a more compact conformation. Surprisingly, there is no intermediate state between the low and high FRET states. Since each NC can bind to 6~8 nucleotides (2), the $(TG)_{10}$ should be able to interact with multiple NC molecules (~3 or 4) at saturating condition. As we mentioned earlier, repeating TG sequence shows high binding affinity to NC, therefore this apparent two state binding manner could be explained that multiple NC bind to $(TG)_{10}$ at the same time. Binding of NC to $(TG)_{10}$ at low concentration of NC is a much slower reaction compared to the reaction at saturating NC, with pseudo-first-order rate constant $k \sim 10^{-3}\ s^{-1}$.

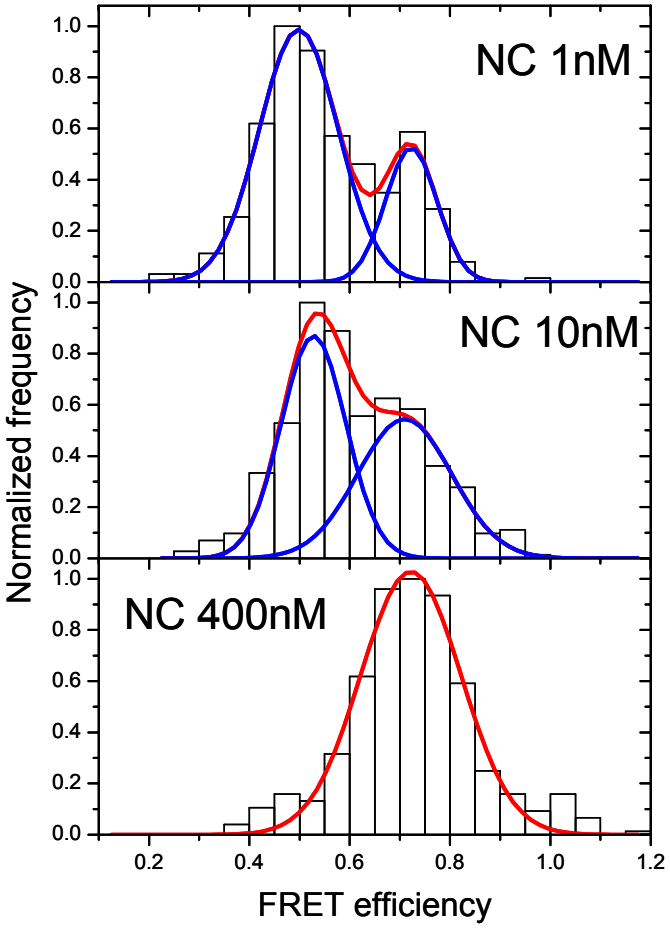


Fig.5.2 Single-molecule FRET efficiency histogram of ssDNA $(TG)_{10}$ in 1nM, 10nM and 400nM NC. The Gaussian fits of the peaks are illustrated in red (overall fit) and blue (individual peak fit).

The dissociation rate of bound NC from the bound complex is limited under buffer only condition, however this dissociation process can be accelerated by introducing high concentration of Mg^{2+} ion, or high concentration of free non-complementary DNA oligonucleotides into the system. After binding reaction of NC and $(TG)_{10}$ reaches equilibrium (Fig. 5.3 stage II), we flowed buffer solution into the reaction

chamber for long time, ~5000s. There was a little amount of molecules with FRET efficiency decreasing to the staring points of 1X buffer only (Fig. 5.3 stage III) as indicated by the fact that main peak of FRET histogram remains the same position after buffer washing. This is different with the case of binding $(TG)_{10}$ by high concentration of Mg^{2+}. For small cation binding to ssDNA, the binding and unbinding reaction is reversible; the dissociation rate is much faster than what we observed for NC (data not shown). This indicates that NC binding and changing conformation of ssDNA is not simply caused by electrostatic interaction within the two. There must be other hydrophobic interaction or hydrogen bond playing important role here. In order to investigate the dissociation process deeply, we flew high concentration of Mg^{2+} after 1X buffer wash (Fig. 5.3 stage IV). The increasing of FRET efficiency here can be explained by the fact that high concentration Mg^{2+} can reduce the stiffness of ssDNA chain more; therefore the average end-to-end distance between two fluorophores was shortened. Last, we used 1X buffer to wash the Mg^{2+} away, giving majority of molecules FRET efficiency back to initiated state (Fig. 5.3 stage VI). Small Mg^{2+} ions penetrate into the NC/ssDNA complex, bind to the negatively charged DNA phosphate backbone and replace some of the bound NC molecules. The accelerating of bound NC dissociation by small cation indicates that electrostatic interaction between NC and ssDNA plays important role in holding the complex.

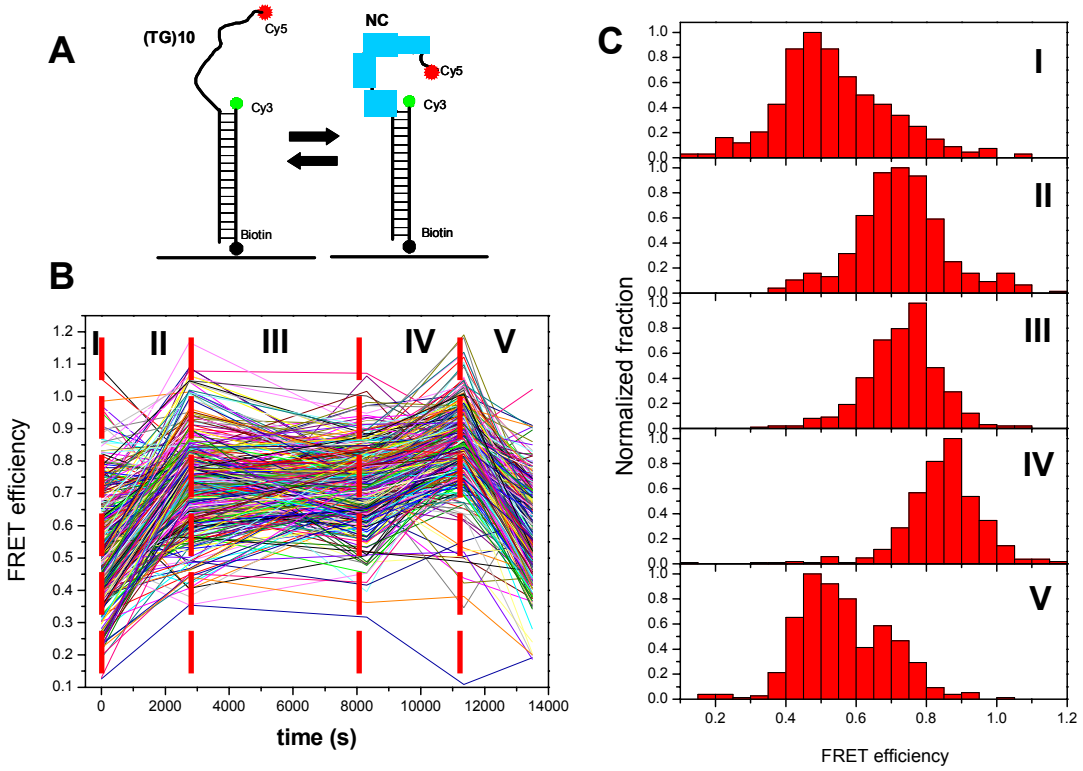


Fig. 5.3 Buffer wash and high Mg^{2+} buffer wash NC bound to ssDNA $(TG)_{10}$
(A) Schematics illustrating the reaction.
(B) Single molecule FRET trajectories. The whole process was divided into five stages: I: buffer; II: 400nM NC; III: buffer wash; IV: buffer containing 20mM Mg2+ wash; V: buffer wash.
(C) Histograms of FRET distribution at different reaction stages.

Next, we tested the DNA sequence effect on the NC binding and induced single-stranded DNA chain conformation change. Oligodeoxythymidylates (oligo-dT_{20}) was employed in the following experiments to keep the same length as $(TG)_{10}$ in previous study. After flowing 400nM NC into the reaction chamber, a shift of FRET distribution toward much higher FRET efficiency was observed. The FRET value at saturated NC condition (0.74±0.06) is comparable to what was observed with $(TG)_{10}$ (0.76±0.09). Followed by a buffer only wash step, a few molecules FRET decreasing to the initiated

state (Fig. 5.4 stage I) leaves majority of the ssDNA molecule in the NC bound form (Fig. 5.4 C). Here each line in Fig. 5.4 B represents a single molecule trajectory; more than 300 molecules trajectories were used to build the FRET histogram in Fig. 5.4 C panel. The replacement washing step used 10uM of free non-complementary short DNA oligos in the buffer solution, the sequence was chosen to have no potential in forming basepairs with ssDNA on the surface. Here we test the idea that whether the bound NC on ssDNA which immobilized on surface will dissociate from the complex and bind onto the free ssDNA oligo in the solution. After a final wash by 1X buffer, ~70% molecules FRET decreased to ~0.5 (Fig. 5.4C stage VI). The same experiment was also performed using $(TG)_{10}$, however due to higher binding affinity between NC and repeated TG sequence compared with polyT, this free non-complementary failed to remove bound NC in this competitive binding reaction.

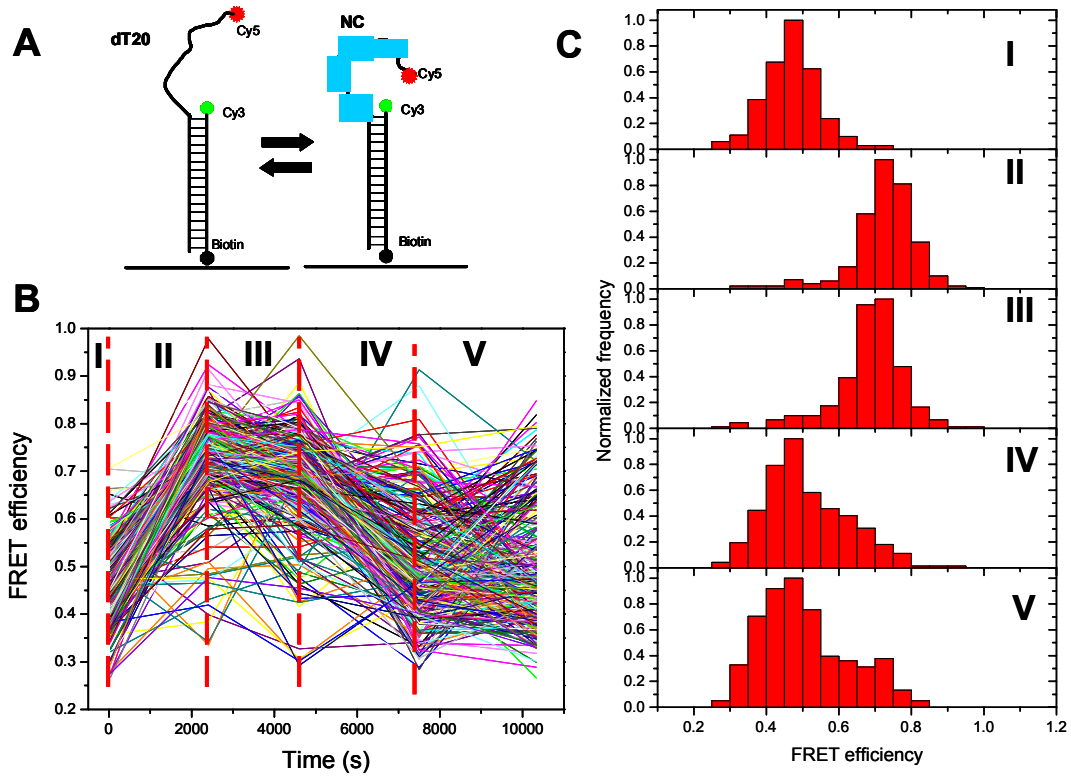


Fig. 5.4 Buffer wash and non-complementary short DNA wash NC bound to ssDNA T_{20}
(A) Schematics illustrating the reaction.
(B) Single molecule FRET trajectories. The whole process was divided into five stages: I: buffer; II: 400nM NC; III: buffer wash; IV: buffer containing 10uM non-complementary 12mer DNA wash (sequence: 5'GCCCGTAAAATTT3'); V: buffer wash.
(C) Histograms of FRET distribution at different reaction stages.

The single-stranded DNA length effect on NC induced conformational changed was also examined using $(TG)_{20}$, which consists 20 repeated TG in the single-stranded region. In 1X buffer condition, the mean FRET efficiency is ~0.2, which is lower than what was observed for $(TG)_{10}$ and polyT_{20}. The average end-to-end distance is increased due to the increasing of total ssDNA length, which is consistent to other reported value (166). 400nM NC can effectively change the single-stranded $(TG)_{20}$ into more compact

conformation, with FRET efficiency risen from ~0.2 to ~0.7 (Fig. 5.5B stage I and II). Again, the 1X buffer wash cannot remove the bound NC away from the $(TG)_{20}$ molecules (Fig. 5.5B stage III). The FRET distribution is very similar to what was observed with 400nM NC in the reaction chamber. Then we flew high concentration of Mg^{2+} (20mM) into buffer solution, FRET efficiency stays high because sufficient amount of Mg^{2+} bind onto single-stranded $(TG)_{20}$ and reduce the chain stiffness by screening of the DNA negative charge. The final step of the 1X buffer wash effectively shifted the FRET distribution to the lower efficiency end. However the FRET distribution is completely different than the distribution of either the initial 1X buffer (Fig. 5.5B stage I) or the 400nM NC (Fig. 5.5B stage II). Two peaks with center values fall in the range of 0.2 to 0.7 are exhibited (Fig. 5.5C bottom panel). This is completely different with the experimental results using $(TG)_{10}$ as we discussed earlier, in which the majority of molecules FRET decreased to the value as in initial 1X buffer. The hypothesis here is this median FRET efficiency we observed for $(TG)_{20}$ is caused by some bound NC molecules that were "trapped" by the long ssDNA chain therefore they were not successfully removed by even high concentration of Mg^{2+} buffer wash. To test this hypothesis, we did the experiments using very low NC concentration, 10nM for $(TG)_{20}$. First we flew the 1X buffer into the reaction chamber followed by 10nM NC, and final wash using the 1X buffer. The FRET distribution did not change after the 1X wash (Fig. 5.5C top). The peak values are ~0.35 and 0.55 respectively, representing two different folded conformations for $(TG)_{20}$ bound with NC. However the FRET distribution with 10nM NC is very similar to that of final wash after 400nM NC (Fig. 5.5B stage V). At very low NC concentration condition, NC will first bind onto the ssDNA where the binding site exhibit higher binding affinity. While at much high NC concentration, 400nM in our experiment, sufficient amount of NC associate with $(TG)_{20}$ molecules, changing their conformation to

a much compact form. After Mg^{2+} competitive binding and washing, some relatively loose bound NC were replaced leaving some stronger bound NC "trapped" within the long ssDNA wrap.

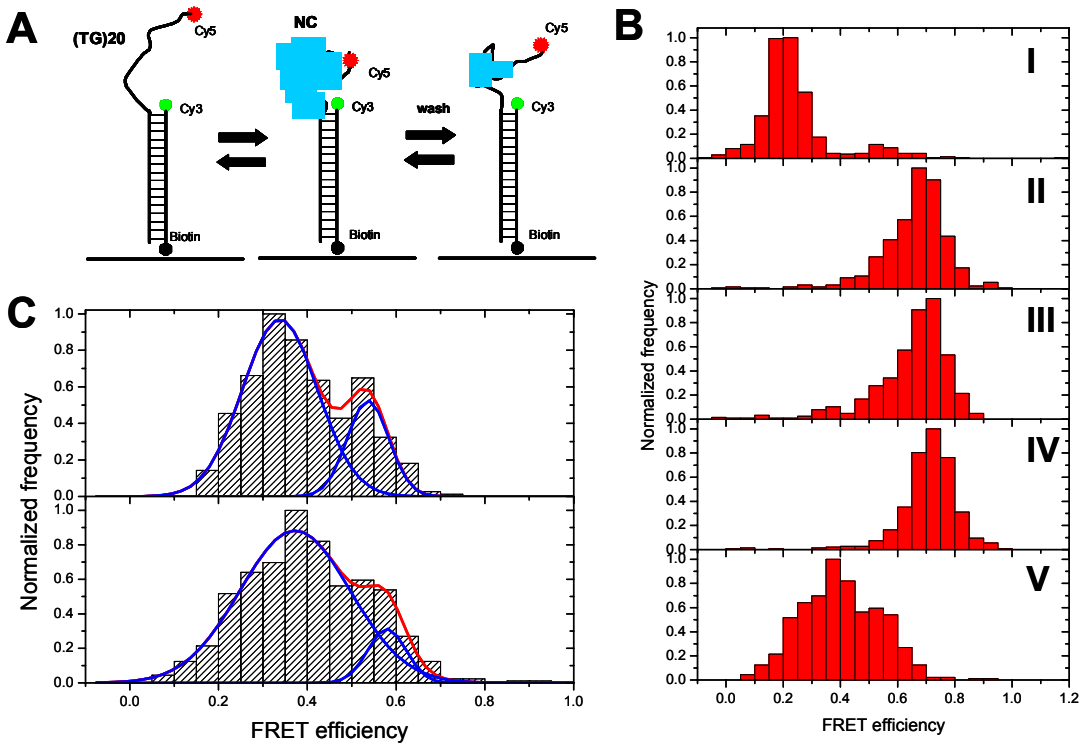


Fig.5.5 Buffer wash and high Mg^{2+} buffer wash NC bound to ssDNA $(TG)_{20}$
(A) Schematics illustrating the reaction.
(B) Histograms of FRET distribution at different reaction stages. I: buffer; II: 400nM NC; III: buffer wash; IV: buffer containing 20mM Mg^{2+} wash; V: buffer wash.
(C) Single-molecule FRET efficiency histograms for (top) buffer wash after flowing 10nM NC and (bottom) reaction stage V shown in (B). The Gaussian fits of the peaks are illustrated in red (overall fit) and blue (individual peak fit).

In order to probe the ssDNA conformation dynamics in the time scale of milliseconds, here we use the individual trajectory mode SM-FRET to investigate

possible fast dynamics existing in this ssDNA/NC complex. In this mode, the laser is focused onto each individual molecule, donor and acceptor intensity are collected at the time with time resolution down to 1ms. Usually more than 25 molecules trajectory are collected in order to build ensemble FRET. This time resolution is much faster than the other mode, scanning image mode, in which the sample is scanned over a 30μm×30μm region and the time resolution for each image is ~300 seconds. Therefore, using this individual trajectory mode, we could probe whether there is any underlying complexity in the ssDNA conformation dynamics which we might miss in the relative slow scanning method. To construct ensemble E_A histograms, the individual molecule trajectories were boxcar time averaged, or smoothed with alternative bin times, τ_B, of 5, 10, and 50 (or 100)ms. In boxcar averaging a group of N adjacent E_A points are averaged together where $N=\tau_B/\tau_D$. Here τ_D is the time spacing in the original non-time-averaged data (1ms). As τ_B is increased the signal-to-noise ratio of the data increases due to averaging. E_A fluctuations are, however, also smoothed by this process, if they occur on a faster time scale than τ_B. Ensemble SM-FRET data for $(TG)_{10}$ in the absence and in the presence of NC is shown in Fig. 5.6. In the absence of NC, the τ_B=5ms E_A histogram shows a single peak. The width of the peak is primarily due to photon shot noise rather than any significant E_A fluctuations, which can also be verify by the small amplitude and fast relaxation of E_A autocorrelation (Fig. 5.7). In the presence of 400nM NC, FRET histogram shows a single peak in the τ_B=5ms, 10ms and 50ms condition respectively. Increasing the boxcar averaging time the peak width shrinks down, which indicates that the broadening is mainly due to photon shot noise. Compare the autocorrelation in the presence of NC and in the absence of NC, we find that the amplitude and the decay rate is on the same order of magnitude. For $(TG)_{10}$, in the presence of NC, no other detectable states is observed within the experimental detection limit. Down to the time limit of a few

millisecond, no other different states for $(TG)_{10}$, /NC complex were observed (see FRET histogram Fig. 5.7).

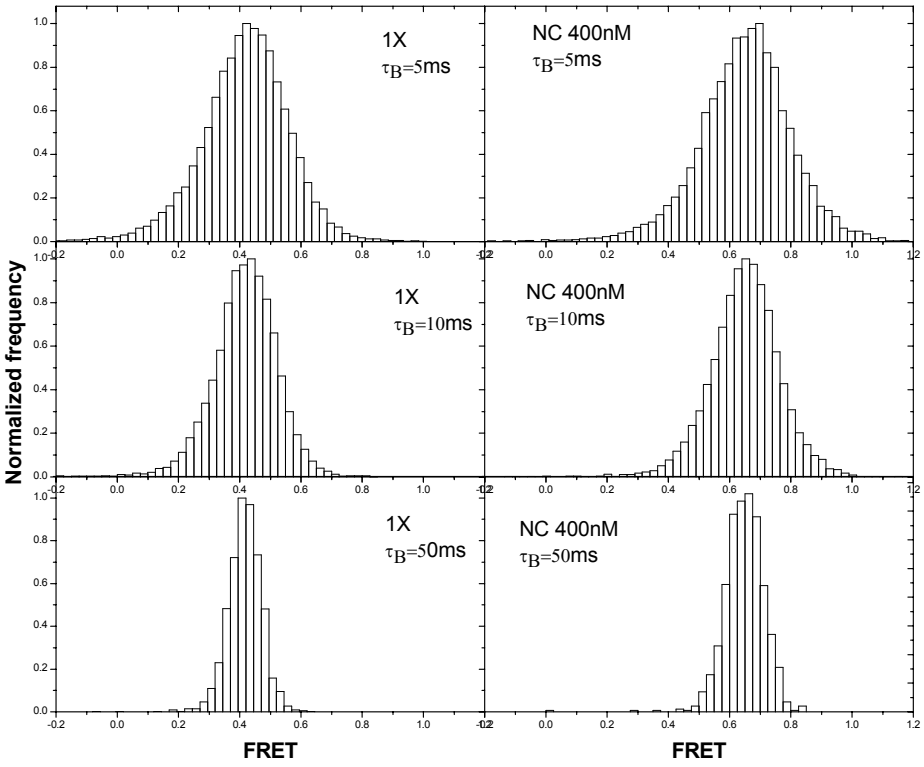


Fig. 5.6. Experimentally obtained ensemble FRET histograms for $(TG)_{10}$ in the absence of NC (left) and in the presence of 400nM NC (right) at three different τ_B.

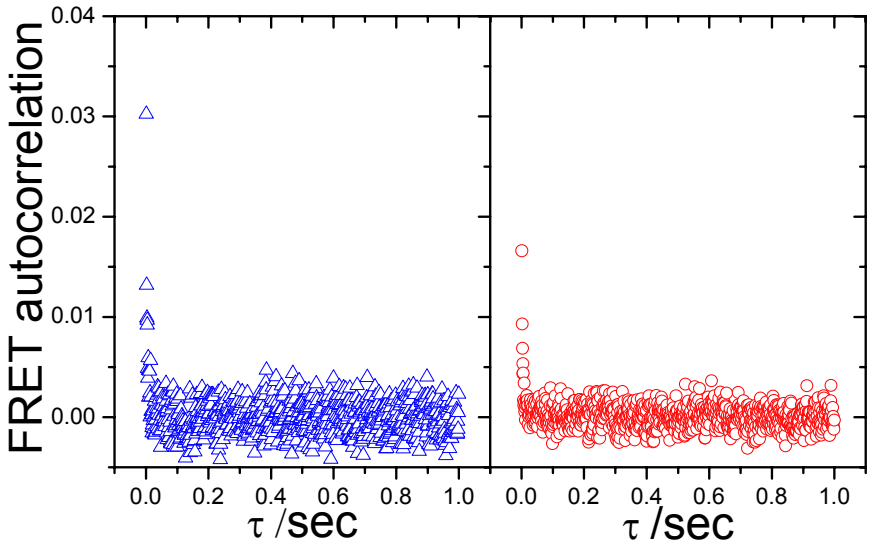


Fig. 5.7. FRET autocorrelation curves for $(TG)_{10}$ in the 1X only (left) and in the presence of 400nM NC. All experiments were performed in the presence of 0.2mM Mg^{2+} and 40mM NaCl.

In order to investigate other simple cation effects on $(TG)_{10}$ FRET distribution in a fast time scale, we performed the FRET measurement using a series of concentrations of Mg^{2+} in the buffer solution under individual trajectory mode. Ensemble E_A histogram with different bin time, τ_B= 5, 10, and 50 (or 100) ms are shown in Fig.5. 8. The trend in average E_A vs [Mg^{2+}] demonstrates clear correlation between $(TG)_{10}$ conformation and positive ion concentration. This is consistent with other reported results of ssDNA with various concentration of Na^+ (166). The persistence length of $(TG)_{10}$ is effectively shortened by increasing Mg^{2+} concentration. The positive ions bind onto ssDNA due to electrostatic interaction and therefore, the average end-to-end distance is reduced which can be predicted by the worm-like-chain model (166). For all Mg^{2+} concentrations, a well shaped single peak was observed in the FRET histogram using various bin time. The end-

to-end distance fluctuation relaxed very fast in our experimental time resolution, thus no other different FRET states is observed for $(TG)_{10}$ under all concentrations of Mg^{2+}. The conformation of $(TG)_{10}$ is in its equilibrium and exhibit a relative 'static' single state end-to-end distance.

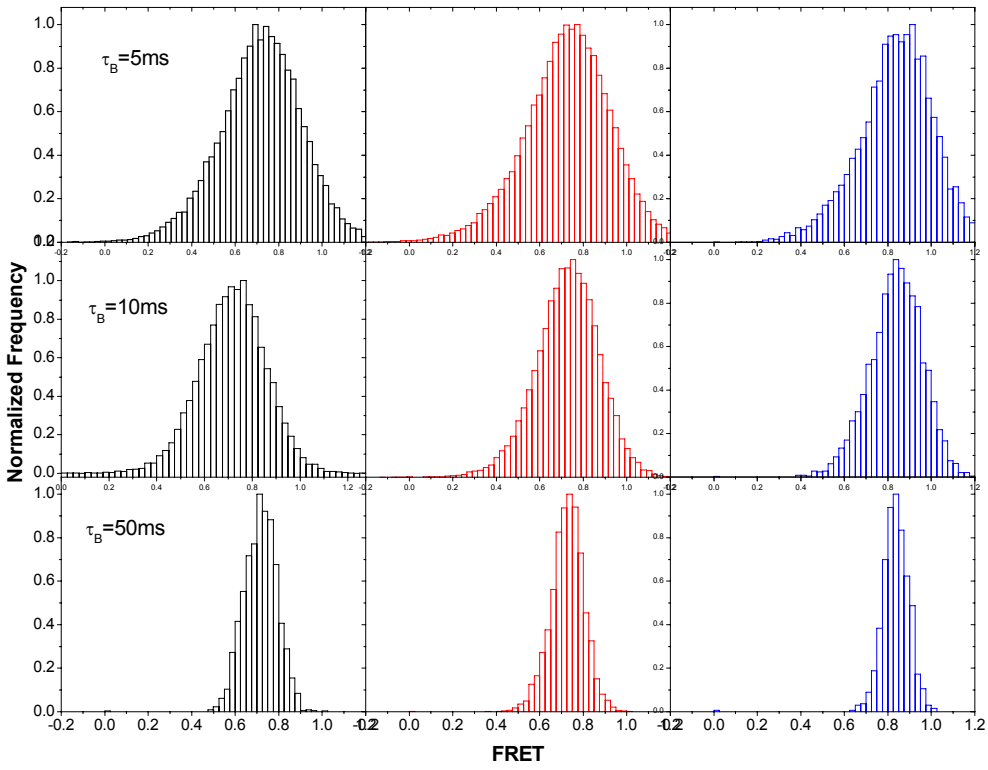


Fig. 5.8 Ensemble FRET histogram for $(TG)_{10}$ at various $[Mg^{2+}]$ concentrations with different τ_B (5ms, 10ms, and 50ms from top to bottom) are compared. On the graph, $[Mg^{2+}]$=2mM (left panel, black); $[Mg^{2+}]$=5mM (middle panel, red); $[Mg^{2+}]$=20mM (right panel, blue)

NC does behave differently than a simple polycation in changing ssDNA conformation. This can be proven by the control experiment using poly-K-Lysine (MW>15,000). 400nM of polylysine was added into buffer solution and flew into

reaction chamber where $(TG)_{10}$ was immobilized on the surface. A significant change in FRET was observed with the peak of E_A shifting from ~0.5 to ~0.85. The FRET efficiency in the presence of 400nM polylysine is higher than that in the presence of NC at the same concentration. This might be due to a more positive charge polylysine carries than that for NC. 1X buffer can wash ~30% of polylysine molecules away within ~4000s (Fig. 5.9). This fast dissociation is significantly different for NC which shows no apparent dissociation using 1X buffer. Although both NC and polylysine can promote the FRET efficiency, which suggests that electrostatic interaction between the ssDNA and NC/polylysine is major component in changing the ssDNA conformation, however, the dissociation rate is quite different for NC and polylysine which indicates there are other important interactions like hydrophobic interaction, H-bond holding the complex.

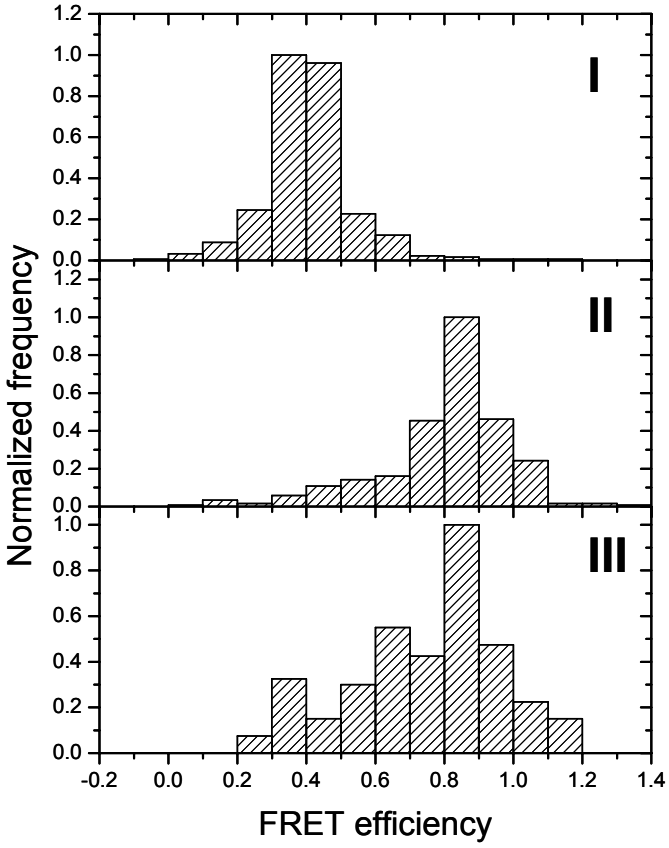


Fig. 5.9. Single-molecule FRET efficiency histogram for $(TG)_{10}$ with poly-L-Lysine and buffer wash. Reaction stages are divided as following. I: buffer; II: 400nM poly-L-Lysine (MW>15,000) reacted for 30min; III: buffer wash after ~4000s.

The zinc finger is known to be of great importance for NC function. Disturbing the zinc finger usually decreases NC activity in chaperone DNA/DNA, DNA/RNA annealing, and other aspects of NC function (2). In order to explore the role of zinc fingers to the interaction of NC and $(TG)_{10}$, we performed the binding experiment in the presence of 1mM EDTA (Fig. 5.10). Ethylenediaminetetraacetic acid (EDTA) is a well know zinc chelating reagent which binds to zinc very strongly and therefore is used

widely in destroying the zinc finger structure of NC (113, 155). 1mM EDTA and 400nM of NC was incubated for 30 minutes before the reaction to complete the zinc ejection process. However, the results we obtained shows that in addition of 1mM EDTA NC still bind effectively to $(TG)_{10}$. 1X buffer washing cannot remove NC from the bound complex which is similar to what we observed for wt NC. This data implies that the zinc fingers are not as crucial as what we expected for the high affinity interaction of NC with $(TG)_{10}$. It has been reported that the basic residues in the N-terminus of NC are critical both for the high affinity of wt NC with $(TG)_4$ and its ability to interact with the bound NC: $(TG)_4$ complex (94). Although EDTA destroyed the zinc finger structure, the whole NC with its N-terminus and basic residues still shows high affinity interaction with $(TG)_{10}$. On the other hand, repeated d(TG) sequence has been shown to have a strong binding affinity with NC (94). Since we expect more than one copy of NC to bind on the $(TG)_{10}$ molecule, the interaction between proteins and the interaction of multiple NC and ssDNA might play a very important role in forming the complex. Additionally, NC still posses its positive charges in the absence of a proper folded zinc-finger. It has been that NC can fold into different conformations to accommodate productive electrostatic interactions with the RNA substrate even in the absence of Zn^{2+} binding (155). The observed shortening of average end-to-end distance is a reflection of the overall more compact conformation compared to naked ssDNA. Within this ssDNA:NC complex, the interaction involving multiple NC and $(TG)_{10}$ might overcome the loss of the zinc finger structure.

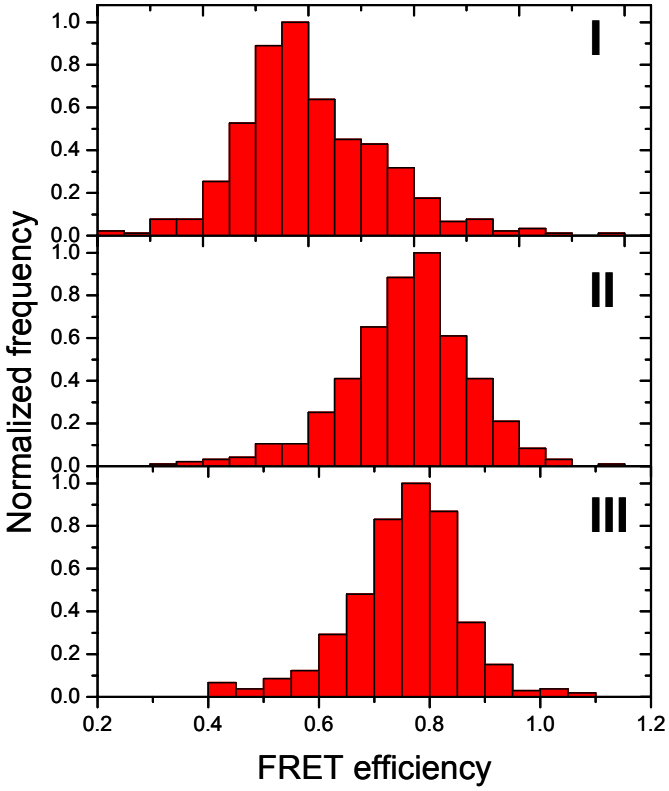


Fig. 5.10 Single-molecule FRET efficiency histogram for $(TG)_{10}$ with EDTA-treated NC and buffer wash. I: buffer; II: EDTA-treated NC (400nM NC was incubated with 1mM EDTA for 30min before reaction); III: buffer wash.

TABLE 5.2 FRET measurements for $(TG)_{10}$, T_{20} and $(TG)_{20}$ in buffer only, 400nM NC and buffer containing 20mM Mg^{2+}

	Buffer	[NC]=400nM	[Mg^{2+}] =20mM
$(TG)_{10}$	0.50±0.08	0.72±0.09	0.86±0.07
T_{20}	0.47±0.06	0.74±0.06	
$(TG)_{20}$	0.20±0.06	0.68±0.08	0.69±0.08

The error is calculated by the width of the fitted Gaussian function for each FRET efficiency histogram.

CONCLUSION

Here we present observations of the NC binding to ssDNA, including d(TG)n and d(T)n, as well as its effect on flexibility and conformation of the oligonucleotide chains. Single-molecule fluorescence resonance energy transfer was used to study NC effect on ssDNA conformation. The K_d for binding of a single $d(TG)_{10}$ molecule to NC is ~10nM. Our results reveal that the rigidity of ssDNA chain is dramatically reduced through interaction with NC. Electrostatic interaction between basic NC and negatively charged ssDNA is a large component of the overall condensation. However, the surprisingly slow dissociation rate of the bound NC under buffer condition and the fact that zinc-finger deficient NC is capable of changing ssDNA conformation, indicate the interaction between these two is even more complex. Dissociation can be accelerated by introducing high concentration of Mg^{2+} buffer or non-complementary oligonucleotide into the system. Finally, our data of long ssDNA chain, $d(TG)_{20}$ unravels heterogeneity of conformational change due to NC binding.

References

1. Karn, J., *Tackling Tat.* Journal of Molecular Biology, 1999. **293**(2): p. 235-254.
2. Levin, J.G., et al., *Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism.* Progress in Nucleic Acid Research and Molecular Biology,, 2005. **80**: p. 217-286.
3. Turner, B.G. and M.F. Summers, *Structural biology of HIV.* J. Mol. Biol., 1999. **285**: p. 1.
4. Myong, S., et al., *Repetitive shuttling of a motor protein on DNA.* Nature, 2005. **437**(7063): p. 1321-1325.
5. Schuler, B., E.A. Lipman, and W.A. Eaton, *Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy.* Nature, 2002. **419**(6908): p. 743-747.
6. Weiss, S., *Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy.* Nature Struct. Biol., 2000. **7**(9): p. 724-729.
7. Xie, Z., et al., *Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme.* Proc. Natl. Acad. Sci., 2004. **101**(2): p. 534-539.
8. Zhuang, X., et al., *Fluorescence Quenching: a Tool for Single-Molecule Protein-Folding Study.* Proc. Natl. Acad. Sci., 2000. **97**(26): p. 14241-14244.
9. Zhuang, X. and M. Rief, *Single-molecule folding.* Curr. Opin. Struct. Biol., 2003. **13**: p. 88-97.
10. Zhuang, X.W., et al., *A single-molecule study of RNA catalysis and folding.* Science, 2000. **288**: p. 2048-2051.
11. Zhuang, X.W., et al., *Correlating structural dynamics and function in single ribozyme molecules.* Science, 2002. **296**(5572): p. 1473-1476.
12. Cosa, G., et al., *Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein.* Biophysical Journal, 2004. **87**(4): p. 2759-2767.
13. Cosa, G., et al., *Evidence for Non-Two-State Kinetics in the Nucleocapsid Protein Chaperoned Opening of DNA Hairpins.* J. Phys. Chem. B, 2006. **110**(5): p. 2419-2426.
14. Guo, J., et al., *Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA.* Journal of Virology, 1997. **71**(7): p. 5178-88.
15. Guo, J., et al., *Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate effecient minus- and plus-strand transfer.* J. Virol., 2000. **74**(19): p. 8980-8988.

16. Liu, H.W., et al., *Single-molecule FRET studies of important intermediates in the nucleocapsid protein chaperoned minus-strand transfer step in HIV-1 reverse transcription.* Biophys. J., 2005. **89**(5): p. 3470-3479.
17. Liu, H.W., et al., *Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(13): p. 5261-5267.
18. Vo, M., I. Rouzina, and K. Musier-Forsyth, *Kinetic mechanizm of the mini-TAR RNA-DNA annealing and the chaperone role of HIV-1 NC protein.* 2004.
19. Vo, M.-N., et al., *Mechanistic Studies of Mini-TAR RNA/DNA Annealing in the Absence and Presence of HIV-1 Nucleocapsid Protein.* Journal of Molecular Biology, 2006. **363**(1): p. 244-261.
20. Zeng, Y.N., et al., *Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(31): p. 12651-12656.
21. Vo, M.N., et al., *HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop-Loop "Kissing" to "Zipper".* Journal of Molecular Biology, 2009. **386**(3): p. 789-801.
22. Bernacchi, S., et al., *HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence.* J. Mol. Biol., 2002. **317**(3): p. 385-399.
23. Lapadat-Tapolsky, M., et al., *Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle.* Nucleic Acids Research, 1993. **21**(4): p. 831-9.
24. Azoulay, J., et al., *Destabilization of the HIV-1 Complementary Sequence of TAR by the Nucleocapsid Protein Through Activation of Conformational Fluctuations.* Journal of Molecular Biology, 2003. **326**: p. 691.
25. Rein, A., L.E. Henderson, and J.G. Levin, *Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication.* Trends in Biochemical Sciences, 1998. **23**(8): p. 297-301.
26. Dibhajj, F., R. Khan, and D.P. Giedroc, *Retroviral Nucleocapsid Proteins Posses Potent Nucleic-Acid Strand Renaturation Activity.* Protein Science, 1993. **2**(2): p. 231-243.
27. Le Cam, E., et al., *Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: An electron microscopy investigation.* Biopolymers, 1998. **45**(3): p. 217-229.
28. Stoylov, S.P., et al., *Aggregation of polyA-HIV-1 nucleocapsid protein NCp7 complexes and properties of the aggregates.* Colloids and Surfaces, A: Physicochemical and Engineering Aspects, 1999. **152**(3): p. 263-274.
29. Stoylov, S.P., et al., *Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein.* Biopolymers, 1997. **41**(3): p. 301-312.

30. Ha, T., et al., *Initiation and reinitiation of DNA unwinding by the E. Coli Rep helicase.* Nature, 2002. **419**: p. 638-641.
31. Kim, H.D., et al., *Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules.* Proc. Natl. Acad. Sci. U.S.A., 2002. **99**: p. 4284-4289.
32. Clegg, R.M., *Fluorescence resonance energy transfer and nucleic acids.* Methods Enzymol, 1992. **211**: p. 353-388.
33. Stryer, L., *Fluorescence energy transfer as a spectroscopic ruler.* Ann. Rev. Biochem., 1978. **47**.
34. Hess, S.T., et al., *Biological and Chemical Applications of Flurescence Correlation Spectroscopy: A Review.* Biochemistry, 2002. **41**(3): p. 697-705.
35. Rasnik, I., S.A. McKinney, and T. Ha, *Nonblinking and longlasting single-molecule fluorescence imaging.* Nature Methods, 2006. **3**(11): p. 891-893.
36. Rasnik, I., et al., *DNA-binding Orientation and Domain Conformation of the E. coli Rep Helicase Monomer Bound to a Partial Duplex Junction: Single-molecule Studies of Fluorescently Labeled Enzymes.* J. Mol. Biol., 2004. **336**(2): p. 395-408.
37. Macklin, J.J., et al., *Imaging and time-resolved spectroscopy of single molecules at an interface.* Science, 1996. **272**(5259): p. 255-258.
38. Xie, S., *Single-molecule Approach to Enzymology.* Single Mol., 2001. **2**(4): p. 229.
39. Xie, X.S. and J.K. Trautman, *Optical Studies of Single Molecules at Room Temperature.* Annu. Rev. Phys. Chem., 1998. **49**: p. 441-480.
40. Ha, T., *Single-molecule Fluorescence Resonance Energy Transfer.* Methods, 2001. **25**: p. 78-86.
41. Selvin, P.R., *The renaissance of fluorescence resonance energy transfer.* Nature Structural Biology, 2000. **7**(9): p. 730-734.
42. Ha, T., et al., *Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(13): p. 6264-6268.
43. Gell, C., D. Brockwell, and A. Smith, *Handbook of Single Molecule Fluorescence Spectroscopy*. 2006: Oxford University Press.
44. Gilboa, E., et al., *A detailed model of reverse transcription and tests of crucial aspects.* Cell, 1979. **18**(1): p. 93-100.
45. Peliska, J.A. and S.J. Benkovic, *Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase.* Science, 1992. **258**(5085): p. 1112-8.
46. Darlix, J.L., et al., *First Glimpses at Structure-Function-Relationships of the Nucleocapsid Protein of Retroviruses.* J. Mol. Biol., 1995. **254**(4): p. 523-537.
47. Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. **270**(36): p. 20871-4.
48. Tsuchihashi, Z. and P.O. Brown, *DNA Strand Exchange and Selective DNA Annealing Promoted by the Human Immunodeficiency Virus Type 1 Nucleocapsid Protein.* J. Virol., 1994. **68**(9): p. 5863-5870.

49. Beltz, H., et al., *Impact of the Terminal Bulges of HIV-1 cTAR DNA on its Stability and the Destabilizing Activity of the Nucleocapsid Protein NCp7.* Journal of Molecular Biology, 2003. **328**(1): p. 95-108.
50. Chang, K.Y. and I. Tinoco, *The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement.* Journal of Molecular Biology, 1997. **269**(1): p. 52-66.
51. Fisher, R.J., et al., *Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides.* Journal of Virology, 1998. **72**(3): p. 1902-1909.
52. Haddrick, M., et al., *Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1.* Journal of Molecular Biology, 1996. **259**(1): p. 58-68.
53. Paillart, J.C., et al., *A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(11): p. 5572-5577.
54. Paillart, J.C., et al., *Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA.* Journal of Molecular Biology, 1997. **270**(1): p. 36-49.
55. Karpel, R.L., L.E. Henderson, and S. Oroszlan, *Interactions of retroviral structural proteins with single-stranded nucleic acids.* J Biol Chem, 1987. **262**(11): p. 4961-7.
56. Urbaneja, M.A., et al., *Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function.* J Mol Biol, 1999. **287**(1): p. 59-75.
57. Godet, J., et al., *During the Early Phase of HIV-1 DNA Synthesis, Nucleocapsid Protein Directs Hybridization of the TAR Complementary Sequences via the Ends of their Double-stranded Stem.* Journal of Molecular Biology, 2006. **356**(5): p. 1180-1192.
58. Vuilleumier, C., et al., *Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study.* Biochemistry, 1999. **38**(51): p. 16816-25.
59. Paoletti, A.C., et al., *Affinities of the Nucleocapsid Protein for Variants of SL3 RNA in HIV-1.* Biochemistry, 2002. **41**(51): p. 15423-15428.
60. Maki, A.H., et al., *Phosphorescence and optically detected magnetic resonance of HIV-1 nucleocapsid protein complexes with stem-loop sequences of the genomic Psi-recognition element.* Biochemistry, 2001. **40**(5): p. 1403-12.
61. McKinney, S.A., et al., *Structural dynamics of individual Holliday junctions.* Nature Struct. Biol., 2003. **10**(2): p. 93-97.
62. Russell, R., et al., *Exploring the folding landscape of a structured RNA.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(1): p. 155-160.
63. Russell, R. and D. Herschlag, *Probing the Folding Landscape of the Tetrahymena Ribozyme: Commitment to Form the Native Conformation is Late in the Folding Pathway.* Journal of Molecular Biology, 2001. **308**(5): p. 839-851.

64. Liu, H.-W., et al., New Insights on the Role of Nucleic Acid/Protein Interactions in Chaperoned Nucleic Acid Rearrangements of HIV-1 Reverse Transcription. **submitted to Proc Natl Acad Sci USA**.
65. Liu, H.-W., et al., *Single-molecule FRET studies of important intermediates in the nucleocapsid protein chaperoned minus-strand transfer step in HIV-1 reverse transcription.* Biophysical Journal, 2005. **89**(5): p. 3470-3479.
66. Green, L.M. and J.M. Berg, *Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants.* Proc. Natl. Acad. Sci. U. S. A., 1990. **87**(16): p. 6403-7.
67. Henderson, L.E., et al., *Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses.* J Biol Chem, 1981. **256**(16): p. 8400-6.
68. Berg, J.M., *Potential metal-binding domains in nucleic acid binding proteins.* Science, 1986. **232**(4749): p. 485-7.
69. Covey, S.N., *Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosiac virus.* Nucleic Acids Res, 1986. **14**(2): p. 623-633.
70. Darlix, J.-L., et al., *Cis Elements and Trans-acting Factors Involved in the RNA Dimerization of the Human Immunodeficiency Virus HIV-1.* J Mol Biol, 1990. **216**: p. 689-699.
71. Feng, Y.-X., et al., *HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro.* Proc. Natl. Acad. Sci., 1996. **93**: p. 7577-7581.
72. Baba, S., et al., *Role of the zinc fingers of HIV-1 nucleocapsid protein in maturation of genomic RNA.* Journal of Biochemistry (Tokyo, Japan), 2003. **134**(5): p. 637-639.
73. Sakaguchi, K., et al., *Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein.* Proc Natl Acad Sci U S A, 1993. **90**(11): p. 5219-23.
74. Fu, W., R.J. Gorelick, and A. Rein, *Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions.* J Virol, 1994. **68**(8): p. 5013-8.
75. Muriaux, D., et al., *NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer.* J. Biol. Chem., 1996. **271**(52): p. 33686-92.
76. Berkowitz, R., J. Fisher, and S.P. Goff, *RNA packaging.* Curr Top Microbiol Immunol, 1996. **214**: p. 177-218.
77. de Rocquigny, H., et al., *Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers.* Proc. Natl. Acad. Sci. USA, 1992. **89**: p. 6472-6476.
78. Prats, A.C., et al., *Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA.* EMBO J., 1988. **7**(6): p. 1777-83.

79. Lapadat-Tapolsky, M., et al., *Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1.* Nucleic Acids Res., 1995. **23**(13): p. 2434-2441.
80. Chan, B., et al., *Intra-tRNA distance measurements for nucleocapsid protein dependent tRNA unwinding during priming of HIV reverse transcription.* Proc. Natl. Acad. Sci., 1999. **96**: p. 459.
81. Feng, Y.X., et al., *The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site.* J. Virol., 1999. **73**(5): p. 4251-6.
82. Darlix, J.-L., et al., *First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses.* J. Mol. Biol., 1995. **254**(4): p. 523-37.
83. Muthuswami, R., et al., *The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence.* J Mol Biol, 2002. **315**(3): p. 311-23.
84. Johnson, P.E., et al., *A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription.* Biochemistry, 2000. **39**(31): p. 9084-91.
85. Wu, T., et al., *Molecular Requirements for Human Immunodeficiency Virus Type 1 Plus-Strand Transfer: Analysis in Reconstituted and Endogenous Reverse Transcription Systems.* J. Virol., 1999. **73**: p. 4794-4805.
86. Auxilien, S., et al., *Role of Post-transcriptional Modifications of Primer tRNALys,3 in the Fidelity and Efficacy of Plus Strand DNA Transfer During HIV-1 Reverse Transcription.* J. Biol. Chem., 1999. **274**: p. 4412-4420.
87. Zhang, Y. and E. Barklis, *Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation.* J. Virol., 1995. **69**(9): p. 5716-5722.
88. Gorelick, R.J., et al., *Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a "zinc finger-like" protein sequence.* Proc. Natl. Acad. Sci. USA, 1988. **85**: p. 8420-8424.
89. Gorelick, R.J., et al., *Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences.* Virology, 1999. **256**(1): p. 92-104.
90. Dorfman, T., et al., *Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein.* J Virol, 1993. **67**(10): p. 6159-6169.
91. Mizuno, A., et al., *Mutational analysis of two zinc finger motifs in HIV type 1 nucleocapsid proteins: effects on proteolytic processing of Gag precursors and particle formation.* AIDS Res Hum Retroviruses, 1996. **12**(9): p. 793-800.
92. Méric, C. and P.F. Spahr, *Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging.* J. Virol., 1986. **60**(2): p. 450-9.
93. de Guzman, R.N., et al., *Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element.* Science, 1998. **279**: p. 384-388.

94. Fisher, R.J., et al., *Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides.* Nucleic Acids Research, 2006. **34**(2): p. 472-484.
95. Luo, G.X. and J. Taylor, *Template switching by reverse transcriptase during DNA synthesis.* J Virol, 1990. **64**(9): p. 4321-8.
96. Berkhout, B. and K.-T. Jeang, *Detailed mutational analysis of TAR RNA: critical spacing between the bulge and loop recognition domains.* Nucleic Acids Res., 1991. **19**(22): p. 6169-6176.
97. Baudin, F., et al., *Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains.* J. Mol. Biol., 1993. **229**(2): p. 382-397.
98. You, J.C. and C.S. McHenry, *Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription.* J. Biol. Chem, 1994. **269**(50): p. 31491-5.
99. Davis, W.R., et al., *Actinomycin D inhibition of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase and nucleocapsid protein.* Biochemistry, 1998. **37**(40): p. 14213-21.
100. Hong, M.K., et al., *Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer.* J. Mol. Biol., 2003. **325**(1): p. 1-10.
101. Fried, M. and D.M. Crothers, *Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis.* Nucleic Acids Res FIELD Full Journal Title:Nucleic acids research FIELD Publication Date:1981. **9**(23): p. 6505-25. FIELD Reference Number: FIELD Journal Code:0411011 FIELD Call Number:.
102. Garner, M.M. and A. Revzin, *A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system.* Nucleic Acids Research, 1981. **9**(13): p. 3047-60.
103. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucl. Acids. Res., 2003. **31**(13): p. 3406-3415.
104. Liu, H.-W., et al., *Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(13): p. 5261-5267.
105. Zeng, Y., et al., *Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(31): p. 12651-12656.
106. Landes, C.F., et al., *Single-Molecule Study of the Inhibition of HIV-1 Transactivation Response Region DNA/DNA Annealing by Argininamide.* Journal of the American Chemical Society, 2007. **129**(33): p. 10181-10188.
107. Harada, Y., et al., *Mechanochemical coupling in actomyosin energy transduction studied by in vitro movement assay.* J. Mol. Biol., 1990. **216**(1): p. 49-68.

108. Kanevsky, I., et al., *Specific Interactions Between HIV-1 Nucleocapsid Protein and the TAR Element.* Journal of Molecular Biology, 2005. **348**(5): p. 1059-1077.
109. Lee, N., R.J. Gorelick, and K. Musier-Forsyth, *Zinc finger-dependent HIV-1 nucleocapsid protein-TAR RNA interactions.* Nucleic Acids Research, 2003. **31**(16): p. 4847-4855.
110. Urbaneja, M.A., et al., *HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity.* Journal of Molecular Biology, 2002. **318**(3): p. 749-764.
111. Mély, Y., et al., *Binding of the HIV-1 nucleocapsid protein to the primer $tRNA_3^{Lys}$ in vitro, is essentially not specific.* J. Biol. Chem., 1995. **270**(4): p. 1650-6.
112. Bernacchi, S., et al., *HIV-1 Nucleocapsid Protein Activates Transient Melting of Least Stable Parts of the Secondary Structure of TAR and its Complementary Sequence.* J. Mol. Biol., 2002. **317**: p. 385-399.
113. Nyborg, J.K. and O.B. Peersen, *That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators.* Biochemical Journal, 2004. **381**(3): p. e3-e4.
114. Bombarda, E., et al., *Molecular mechanism of the Zn2+-induced folding of the distal CCHC finger motif of the HIV-1 nucleocapsid protein.* Biophysical Journal, 2007. **93**(1): p. 208-217.
115. Levin, J.G., et al., *Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism.* Prog Nucleic Acid Res Mol Biol, 2005. **80**: p. 217-286.
116. Summers, M.F., et al., *High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach.* Biochemistry, 1990. **29**(2): p. 329-40.
117. Cruceanu, M., et al., *Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins.* Nucleic Acids Research, 2006. **34**(2): p. 593-605.
118. Beltz, H., et al., *Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7.* J. Mol. Biol., 2004. **338**(4): p. 711-23.
119. Surovoy, A., et al., *Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1.* Journal of Molecular Biology, 1993. **229**(1): p. 94-104.
120. Zuker, M., D.H. Mathews, and D.H. Turner, *Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide*. RNA Biochemistry and Biotechnology, ed. J. Barciszewski and B.F.C. Clark. 1999, Dordrecht: Kluwer Academic Publishers. 11-43.
121. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(4): p. 1460-1465.
122. Zhuang, X.W., et al., *A single-molecule study of RNA catalysis and folding.* Science, 2000. **288**(5473): p. 2048.

123. Weiss, S., *Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy.* Natural Structural Biology, 2000. **7**(9): p. 724-729.
124. Xie, Z., et al., *Single-molecule studies highlight conformational heterogeneity in the early folding steps of a large ribozyme.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(2): p. 534-539.
125. Munro, J.B., et al., *Identification of two distinct hybrid state intermediates on the ribosome.* Molecular Cell, 2007. **25**(4): p. 505-517.
126. Fei, J., et al., *Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation.* Molecular Cell, 2008. **30**(3): p. 348-359.
127. Theissen, B., et al., *Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase.* Proc. Natl. Acad. Sci. U.S.A., 2008. **105**(2): p. 548-553.
128. Feng, S. and E.C. Holland, *HIV-1 tat trans-activation requires the loop sequence within tar.* Nature, 1988. **334**: p. 165-167.
129. Weeks, K.M., et al., *Fragments of the HIV-1 Tat protein specifically bind TAR RNA.* Science, 1990. **249**: p. 1281-1285.
130. Cordingley, M.G., et al., *Sequence-specific interaction of Tat protein and Tat peptide with the transactivation-responsive sequence element of human immunodeficiency virus type 1 in vitro.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**: p. 8985-8989.
131. Frankel, A.D., *Peptide models of the Tat-TAR protein-RNA interaction.* Protein Science, 1992. **1**: p. 1539-1542.
132. Calnan, B.J., et al., *Arginine-mediated RNA recognition: The Arginie fork.* Science, 1991. **252**(1167-1171).
133. Derse, D., et al., *A minimal lentivirus Tat.* Journal of Virology, 1991. **65**(12): p. 7012-7015.
134. Tan, R.Y., et al., *RNA recognition by nHIV-1 Tat and Rev.* Seminars in Virology, 1997. **8**: p. 186-193.
135. Guo, J.H., et al., *Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer.* Journal of Virology, 2000. **74**(19): p. 8980-8988.
136. Cosa, G., et al., *Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1NC protein.* Biophysical Journal, 2004. **87**(4): p. 2759-2767.
137. You, J.C. and C.S. McHenry, *HUMAN-IMMUNODEFICIENCY-VIRUS NUCLEOCAPSID PROTEIN ACCELERATES STRAND TRANSFER OF THE TERMINALLY REDUNDANT SEQUENCES INVOLVED IN REVERSE TRANSCRIPTION.* Journal of Biological Chemistry, 1994. **269**(50): p. 31491-31495.
138. Liu, H.W., et al., *Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription.* Biophysical Journal, 2005. **89**(5): p. 3470-3479.

139. Cosa, G., et al., *Evidence for non-two-state kinetics in the nucleocapsid protein chaperoned opening of DNA hairpins.* Journal of Physical Chemistry B, 2006. **110**(5): p. 2419-2426.
140. Liu, H.W., et al., *Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(13): p. 5261-5267.
141. Zeng, Y.N., et al., *Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(31): p. 12651-12656.
142. Vo, M.N., et al., *HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop-Loop "Kissing" to "Zipper".* J. Mol. Biol., 2009. **386**(3): p. 789-801.
143. Vo, M.N., et al., *Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein.* J. Mol. Biol., 2006. **363**(1): p. 244-261.
144. Vo, M.N., et al., *Effect of Mg2+ and Na+ on the Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Implications for Reverse Transcription.* J. Mol. Biol., 2009. **386**(3): p. 773-788.
145. Hong, M.K., et al., *Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer.* Journal of Molecular Biology, 2003. **325**(1): p. 1-10.
146. Azoulay, J., et al., *Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations.* Journal of Molecular Biology, 2003. **326**(3): p. 691-700.
147. Lapadat-Tapolsky, M., et al., *NTERACTIONS BETWEEN HIV-1 NUCLEOCAPSID PROTEIN AND VIRAL-DNA MAY HAVE IMPORTANT FUNCTIONS IN THE VIRAL LIFE-CYCLE.* Nucleic Acids Research, 1993. **21**(4): p. 831-839.
148. Dibhajj, F., R. Khan, and D.P. Giedroc, *RETROVIRAL NUCLEOCAPSID PROTEINS POSSES POTENT NUCLEIC-ACID STRAND RENATURATION ACTIVITY.* Protein Science, 1993. **2**(2): p. 231-243.
149. Stoylov, S.P., et al., *Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein.* Biopolymers, 1997. **41**(3): p. 301-312.
150. Landes, C.F., et al., *Single-molecule study of the inhibition of HIV-1 transactivation response region DNA/DNA annealing by argininamide.* Journal of the American Chemical Society, 2007. **129**(33): p. 10181-10188.
151. Zhang, J.L., P.L. Sharma, and C.S. Crumpacker, *Enhancement of the basal-level activity of HIV-1 long terminal repeat by HIV-1 nucleocapsid protein.* Virology, 2000. **268**(2): p. 251-263.

152. Zhang, J.L. and C.S. Crumpacker, *Human immunodeficiency virus type 1 nucleocapsid protein nuclear localization mediates early viral mRNA expression.* Journal of Virology, 2002. **76**(20): p. 10444-10454.
153. Ha, T., *Single-molecule fluorescence methods for the study of nucleic acids.* Current Opinion in Structural Biology, 2001. **11**(3): p. 287-292.
154. Cruceanu, M., et al., *Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins.* Nucleic Acids Res., 2006. **34**(2): p. 593-605.
155. Urbaneja, M.A., et al., *Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: Elucidation of the structural determinants for function.* Journal of Molecular Biology, 1999. **287**(1): p. 59-75.
156. Kameoka, M., et al., *The Tat protein of human immunodeficiency virus type 1 (HIV-1) can promote placement of tRNA primer onto viral RNA and suppress later DNA polymerization in HIV-1 reverse transcription.* Journal of Virology, 2002. **76**(8): p. 3637-3645.
157. Kuciak, M., et al., *The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro.* Nucleic Acids Research, 2008. **36**(10): p. 3389-3400.
158. Hargittai, M.R.S., et al., *HIV-1 nucleocapsid protein zinc finger structures induce tRNA(Lys,3) structural changes but are not critical for primer/template annealing.* Journal of Molecular Biology, 2001. **312**(5): p. 985-997.
159. You, J.C. and C.S. McHenry, *HIV nucleocapsid protein. Expression in Escherichia coli, purification, and characterization.* J. Biol. Chem., 1993. **268**(22): p. 16519-27.
160. Urbaneja, M.A., et al., *HIV-1 nucleocapsid protein as a nucleic acid chaperone: Spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity.* J. Mol. Bio., 2002. **318**: p. 749-764.
161. Hagan, N. and D. Fabris, *Direct Mass Spectrometric Determination of the Stoichiometry and Binding Affinity of the Complexes between Nucleocapsid Protein and RNA Stem-Loop Hairpins of the HIV-1 Y-Recognition Element.* Biochemistry, 2003. **42**(36): p. 10736-10745.
162. Morellet, N., et al., *Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC).* J Mol Biol, 1998. **283**(2): p. 419-34.
163. Amarasinghe, G.K., et al., *NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform.* J Mol Biol, 2000. **299**(1): p. 145-56.
164. Avilov, S.V., et al., *Site-Specific Characterization of HIV-1 Nucleocapsid Protein Binding to Oligonucleotides with Two Binding Sites.* Biochemistry, 2009. **48**(11): p. 2422-2430.
165. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome.* Nature, 2009. **460**(7256): p. 711-U87.
166. Murphy, M.C., et al., *Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy.* Biophysical Journal, 2004. **86**(4): p. 2530-2537.

Vita

Xiaojing Ma was born in Xi'an, Shaanxi, China on September 19th, 1982, the daughter of Baoxia Li and Mengxiong Ma. After graduating from the Middle school affiliated to the Northwestern Polytechnical University, Xi'an, Shaanxi, she attended the University of Science and Technology of China (USTC), Hefei, Anhui, in 2001. During her undergraduate education, she did her undergraduate research with Professor Haojun Liang working on computer simulation of polymer materials. In July 2005, she graduated with a Bachelor of Science in Chemistry. In August 2005, she entered the Department of Chemistry and Biochemistry at the University of Texas at Austin to pursue PhD degree in Chemistry. Under the supervision of Professor Paul F. Barbara, she worked on the project of studying HIV-1 nucleocapsid protein catalytic function in rearrangement of nucleic acid structures. During her studies, in 2005 and 2006, she was awarded the Dr. Morgan scholarship.

Permanent address: 303 Middle Unit, #2 Building, Haojia Chengshi Huayuan, West Yan Ta Road, Xi'an, Shaanxi, 710061, China

This dissertation was typed by Xiaojing Ma